NGL-05-003-016

FROM: "System Theory"
L. A. Zadeh and E. Polak, eds.
McGraw-Hill Book Co. 1969

1

The concepts of system, aggregate, and state in system theory

L. A. Zadeh[1]
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley

1 *Introduction*

What is a system? What is a state of a system? How can these notions be given precise and yet general meaning, and how can they be related to some of the other basic concepts in system theory? These are the central questions to which our introductory chapter is addressed.

A system, according to a dictionary definition, is a collection of objects united by some form of interaction or interdependence. In this sense, almost everything is a system of one kind or another. Indeed, in its broad interpretation, the notion of a system is one of the most pervasive ideas in the domain of human thought.[2]

Not quite so pervasive, but very basic nonetheless, is the notion of a state of a system. Roughly, a state of a system at any given time is the information needed to determine the behavior of the system from that time on. Thus, in kinematics, the state of a rigid body in rectilinear motion is its position and velocity. In electric circuits, the state of a network is a vector whose components are the capacitor voltages and inductor currents. In the case of a thermochemical catalytic process, the state may be the concentrations of the reagents, their temperatures, and their partial pressures.

[1] This research was supported in part by the National Aeronautics and Space Administration under Grant ~~NsG 354~~-(s-5) and in part by the National Science Foundation under Grant GK-2277. NGL-05-003-016

[2] Discussions of the concepts of system and state from modern points of view may be found in [1–9] and their references.

In what follows, we shall not attempt to present a survey or a critique of the many special ways in which the notions of system and state have been defined in various fields of science. Rather, our aim is to define these and related notions in a general and mathematically precise way, with a view to providing a firm foundation for the construction of a broad conceptual framework for system theory. Our point of departure will be the same as in [1–2], that is, the conception of a system as a set of input-output pairs.[1] However, our approach will be simpler and more direct. In order to make our exposition self-contained, we do not assume that the reader is familiar with the material presented in [1–2].[2]

As will be seen later, when one starts with the definition of a system as a set of input-output pairs, a state of the system can be defined very naturally as a tag attached to a subset of input-output pairs satisfying certain consistency conditions. Such a subset will be referred to as an *aggregate*. The concept of an aggregate will play an important role in our approach and, in particular, will serve to provide a very natural way for introducing the basic ideas of state equivalence, system equivalence, input-output-state relations, etc.

Notational preliminaries

For the convenience of the reader, we list first several commonly used abbreviations and conventions which will be employed without further explanation in this chapter.

1. Lowercase boldface letters denote vectors, for example, $\boldsymbol{\alpha}$, $\mathbf{u}$, $\mathbf{x}$.
2. Greek and italic letters denote scalar-valued variables, functions, and operators, for example, α, $u(t)$, L.
3. Braces denote a set or a family, for example, $\{x\}$ is a set with

[1] The definition of a system as a set of input-output pairs and, more particularly, as a *relation* rather than as a mapping or an operator (see *1.2.8*), constitutes one of the basic differences between the point of view introduced in [1] and the conventional definitions of a system which one finds in the literature of control theory and automata theory. In these theories, a system is usually defined through its state equations or as a mapping from the space of inputs to the space of outputs.

[2] The important work of K. Krohn and J. Rhodes [10–11] and subsequent contributions by P. Zeiger [12], R. Kalman [5], M. Arbib [26], T. Windeknecht [9], and others have shown that some of the concepts and techniques of abstract algebra can be useful in analyzing the behavior of certain types of systems, especially finite-state systems and linear discrete-time systems. Even in the case of the very general class of systems considered in this chapter, some of the definitions and results can be expressed more compactly through the use of algebraic notation and terminology. We do not use such notation herein in order not to make familiarity with modern algebra a prerequisite for the understanding of our exposition of the basic concepts of system theory.

generic element x. $\{x|P\}$ is a set of x's having property P, for example, $\{x|5 < x < 10\}$ is the open interval (5,10). When necessary, the notation $\{f(t,\alpha)\}_\alpha$ will be used to place in evidence the variable (α in this case) which generates the family. The terms *set, collection, family,* and *space* will be used interchangeably.

4. The symbol (a,b) is used in two different senses: (1) As an ordered pair of variables a and b, and (2) as the open interval $\{x|a < x < b\}$, when a and b take values on the real line $(-\infty,\infty)$. In the latter case, the symbol $[a,b]$ will be used, as usual, to denote the closed interval $\{x|a \leq x \leq b\}$.

5. $\triangleq$ stands for *equal by definition*, or *denotes*.

6. $\Rightarrow$ stands for *implies*. Thus, $A \Rightarrow B$ means "In order that B be true, it is sufficient that A be true."

7. $\Leftrightarrow$ denotes *implies and is implied by*.

8. $\forall$ is an abbreviation for *for all;* $\exists$ is an abbreviation for *there exists*.

9. $\in$ stands for *belongs to* or *is an element of*. $\notin$ stands for *does not belong to*.

10. R^n denotes the space of ordered n-tuples of real numbers. Thus, $R^1 \triangleq (-\infty,\infty)$ and $R^n \triangleq \{(x_1, \ldots, x_n)\}$, with $x_i \in R^1$, $i = 1, \ldots, n$.

11. If $X = \{x\}$ and $Y = \{y\}$, then $X \times Y \triangleq \{(x,y)\}$.

We turn next to notations which are specific to system theory. The symbol t will stand, as usual, for time. The range of the variable t will be denoted by T, with the understanding that unless otherwise indicated T is the real line $(-\infty,\infty)$. In the case of so-called *discrete-time systems*, T will be assumed to be the set of integers $\ldots, -1, 0, 1, 2, \ldots$.

A time function $\mathbf{v}$ is understood to be a set of pairs $\{(t,\mathbf{v}(t))\}$, $t \in T$, where $\mathbf{v}(t)$ denotes the *value* of $\mathbf{v}$ at time t. The range of $\mathbf{v}(t)$ will be denoted by $R[\mathbf{v}(t)]$ and, unless stated to the contrary, will be assumed to be independent of t. Usually, for each t, $\mathbf{v}(t)$ will be an ordered n-tuple of real numbers, so that $R[\mathbf{v}(t)] = R^n$.

If $\mathbf{v}$ is a time function defined over $(-\infty,\infty)$, then the set of pairs $\{(t,\mathbf{v}(t))\}$, $t \in I$, where I is an interval, is called a *segment* of $\mathbf{v}$ *over the observation interval* I or, for short, a *segment*. Such a segment will be denoted by $\mathbf{v}_I$ or, more explicitly, by $\mathbf{v}_{[t_0,t_1]}$, if $I = [t_0,t_1]$. Correspondingly, we shall write $\mathbf{v}_{(t_0,t_1]}$ when I is the semiclosed interval $(t_0,t_1]$; and $\mathbf{v}_{(t_0,t_1)}$ when I is the open interval (t_0,t_1). When no confusion with the time function $\mathbf{v}$ can arise, its segment $\mathbf{v}_I$ will be abbreviated to $\mathbf{v}$. The length of $\mathbf{v}$ will be denoted by $l(\mathbf{v})$ and will be identified with the length of the interval $[t_0,t_1]$. Thus $l(\mathbf{v}) \triangleq t_1 - t_0$.

A segment of $\mathbf{v}$ which comprises a segment $\mathbf{v}^0$ followed by a segment $\mathbf{v}^1$ will be denoted by $\mathbf{v}^0\mathbf{v}^1$. More explicitly, if $\mathbf{v}^0 \triangleq \mathbf{v}_{[t_0,t_1]}$ and $\mathbf{v}^1 \triangleq \mathbf{v}_{[t_1,t_2]}$, then $\mathbf{v}^0\mathbf{v}^1 \triangleq \mathbf{v}_{[t_0,t_2]}$, with $\mathbf{v}^0$ and $\mathbf{v}^1$ being segments of $\mathbf{v}^0\mathbf{v}^1$. (See Fig. *1.1.1.*) Note that

$$l(\mathbf{v}^0\mathbf{v}^1) = l(\mathbf{v}^0) + l(\mathbf{v}^1)$$

In illustrating various concepts which will be introduced in later sections, it will frequently be convenient to deal with time functions

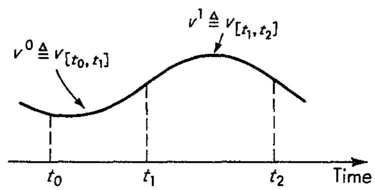


Fig. 1.1.1 Product notation for contiguous segments of a time function.

which are sequences of binary symbols, for example, 0, 1 or a, b. According to the convention stated earlier, such a time function should be written as, say, $v = \{(0,a),(1,b),(2,b),(3,a),(4,a), \ldots\}$, where the first element of each pair indicates the time of occurrence of the second element. To reduce the cumbersomeness of this notation, we shall write v in the usual form, $v = abbaa \cdots$, indicating separately, if needed, the time of occurrence of the first symbol.

This completes the notational preliminaries. We are now ready to turn to defining some of the basic concepts of system theory.

A word of warning: On first exposure, an uninitiated reader may find the definitions formulated in the following sections to be somewhat artificial and bearing little resemblance to his prior notions of the meaning of such terms as system, state, etc. In time, however, he will come to realize that these abstract-sounding definitions are entirely natural and that they merely serve to give precise meaning, couched in mathematical notation and terminology, to various ill-defined notions which occur in the traditional approaches to system analysis.

2 *Definition of an abstract object*

As a preliminary to defining a system, we shall introduce a closely related notion—the notion of an *abstract object*.

In order to avoid stating various definitions both for the case where t is continuous (that is, its range is an interval) and for the case where t is discrete ($t = \ldots, -1,0,1,2, \ldots$), we shall restrict our attention to the former case and assume that T, the range of t, is the real line. With minor modifications, the same definitions will usually apply when t is discrete or, more generally, when t ranges over a subset of the real line.

With this understanding, consider a physical object $\mathcal{P}$ which is associated with a set of measurable attributes, say, mass, height, width, color, velocity, etc. Let the values of these attributes at time t be denoted by $v_1(t), \ldots, v_n(t)$. Clearly, from the mathematical point of view, what matters about $\mathcal{P}$ is not the physical identity of the time functions $v_1, \ldots, v_n$ but the mathematical relations between them. In essence, it is these relations, stripped of their physical identity, that constitute an *abstract object.*

To transform this rather vague idea into a precisely defined concept, it will be helpful to first focus our attention on a consistency condition which will play an important role in subsequent definitions. The significance of this condition will become clearer later, after the notion of an abstract object is related to that of its state.

Closure under segmentation

For each t_0 and t_1 in $(-\infty, \infty)$, with $t_1 \geq t_0$, let $V \triangleq \{\mathbf{v}_{[t_0,t_1]}\}$ denote a family of segments, with the understanding that to each interval $[t_0,t_1]$ may correspond more than one segment $\mathbf{v}_{[t_0,t_1]}$.

Let $\tau \in [t_0,t_1]$ and let the segments $\mathbf{v}_{[t_0,t_1]}$, $\mathbf{v}_{[t_0,\tau]}$, and $\mathbf{v}_{[\tau,t_1]}$ be denoted by $\mathbf{v}$, $\mathbf{v}^0$, and $\mathbf{v}^1$, respectively. Then, in accordance with the notation for contiguous segments, we can write $\mathbf{v} = \mathbf{v}^0\mathbf{v}^1$.

The family of segments $V = \{\mathbf{v}^0\mathbf{v}^1\}$ will be said to be *closed under segmentation* or, for short, *satisfy the CUS condition* if and only if for each τ in $[t_0,t_1]$ both $\mathbf{v}^0$ and $\mathbf{v}^1$ belong to V. In symbols, the CUS condition may be expressed as

1 $$\mathbf{v}^0\mathbf{v}^1 \in V \Rightarrow \mathbf{v}^0 \in V \quad \text{and} \quad \mathbf{v}^1 \in V$$

for all t_0, t_1 in $(-\infty, \infty)$, all $\mathbf{v}$ in V, and all τ in $[t_0,t_1]$. Equivalently, the family V is closed under segmentation if and only if every segment of a segment which is a member of V is also a member of V.

Note that from $\mathbf{v}^0 \in V$ and $\mathbf{v}^1 \in V$ it does not necessarily follow that $\mathbf{v}^0\mathbf{v}^1 \in V$. If $\mathbf{v}^0 \in V$ and $\mathbf{v}^0\mathbf{v}^1 \in V$, then $\mathbf{v}^1$ will be said to be a *continuation* of $\mathbf{v}^0$ in V.

2 *Example* As a simple illustration, let $v_{[t_0,t_1]}$ be defined by the equation

$$v(t) = \alpha + \beta(t - t_0) \qquad t_0 \leq t \leq t_1$$

where t_0, t_1, α, and β range over $(-\infty,\infty)$. Clearly, the CUS condition is satisfied in this case. On the other hand, if $v(t)$ were defined by

$$v(t) = 1 + \beta(t - t_0)$$

then the CUS condition would not be satisfied.

The CUS condition can readily be extended to the case where v is an ordered n-tuple of segments (defined on the same interval) and V is a family of such n-tuples. For example, if $\mathbf{v}$ is an ordered pair, $\mathbf{v} = (\mathbf{u},\mathbf{y})$, where $\mathbf{u} \triangleq \mathbf{u}_{[t_0,t_1]}$ and $\mathbf{y} \triangleq \mathbf{y}_{[t_0,t_1]}$, then the CUS condition *1* becomes

3 $$(\mathbf{u}^0\mathbf{u}^1,\mathbf{y}^0\mathbf{y}^1) \in V \Rightarrow (\mathbf{u}^0,\mathbf{y}^0) \in V \quad \text{and} \quad (\mathbf{u}^1,\mathbf{y}^1) \in V$$

where $\mathbf{u}^0 \triangleq \mathbf{u}_{[t_0,\tau]}$, $\mathbf{u}^1 \triangleq \mathbf{u}_{[\tau,t_1]}$, $\mathbf{y}^0 \triangleq \mathbf{y}_{[t_0,\tau]}$, and $\mathbf{y}^1 \triangleq \mathbf{y}_{[\tau,t_1]}$. As before, a pair $(\mathbf{u}^1,\mathbf{y}^1)$ will be said to be a *continuation* of $(\mathbf{u}^0,\mathbf{y}^0)$ in V if $(\mathbf{u}^0\mathbf{u}^1,\mathbf{y}^0\mathbf{y}^1) \in V$.

Abstract objects

Earlier in this section, an abstract object was vaguely identified with a set of relations among the time functions which represent its attributes. Having formulated the CUS condition, we are now in a position to define the concept of an abstract object in concrete and yet very general terms.

4 **Definition** An abstract object $\mathcal{A}$ is a family of ordered pairs of time functions

$$\mathcal{A} = \{(\mathbf{u}_{[t_0,t_1]},\mathbf{y}_{[t_0,t_1]})\} \qquad t_0,\ t_1 \in (-\infty,\infty)$$

satisfying the CUS condition, that is, $(\mathbf{u}_{[t_0,t_1]},\mathbf{y}_{[t_0,t_1]})$ belongs to $\mathcal{A} \Rightarrow$ every segment of $(\mathbf{u}_{[t_0,t_1]},\mathbf{y}_{[t_0,t_1]})$ belongs to $\mathcal{A}$.

The segments $\mathbf{u} \triangleq \mathbf{u}_{[t_0,t_1]}$ and $\mathbf{y} \triangleq \mathbf{y}_{[t_0,t_1]}$ are termed, respectively, *input* and *output* of $\mathcal{A}$, and the pair $(\mathbf{u},\mathbf{y})$ is said to be an *input-output pair belonging to* $\mathcal{A}$. Thus, in essence, an abstract object $\mathcal{A}$ is a collection of input-output pairs,

5 $$\mathcal{A} = \{(\mathbf{u},\mathbf{y})\}$$

satisfying the CUS condition.

The families of time functions

6 $$\mathcal{D}(\mathcal{A}) \triangleq \{\mathbf{u}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}\}$$

and

7 $$\mathcal{R}(\mathcal{A}) \triangleq \{\mathbf{y}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}\}$$

constitute, respectively, the *domain* and *range* of $\mathcal{A}$. These families represent, respectively, the sets of all input and outputs that can be associated with $\mathcal{A}$. For a diagrammatic representation of an abstract object $\mathfrak{A}$ with input $\mathbf{u}$ and output $\mathbf{y}$, we shall employ the conventional block diagram form shown in Fig. *1.2.1*.

8 *Comment* The above definition of an abstract object is merely a formal expression of the fact that any interaction with a physical object involves varying some of the attributes of this object and observing the resulting variations in other attributes. The attributes that are varied play the role of inputs (causes) and the resulting variations are the outputs (effects). In taking the point of view that an abstract object is a collection of its input-output pairs, we are in effect asserting that so long as our interaction with an object takes

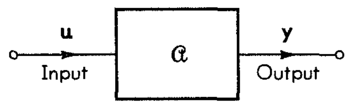


Fig. 1.2.1 Diagrammatic representation of a system $\mathfrak{A}$ with input $\mathbf{u}$ and output $\mathbf{y}$.

place through the observation of the variations in its attributes, it is logical to identify the object in question with the totality of possible observations of variations in its attributes. The totality of such observations, then, constitutes an abstract object.

It is important to observe that the definition of an abstract object does not imply that to each input $\mathbf{u}$ corresponds a unique output $\mathbf{y}$. On the contrary, to each input $\mathbf{u}$ will correspond in general a number of possible outputs, each of which constitutes a possible *response* of $\mathfrak{A}$ to $\mathbf{u}$. As will be seen later, the nonuniqueness of response to a given input reflects the dependence of the output not only on the input but also on the initial state of the object. In mathematical terminology, because of the nonuniqueness of the dependence of output on input, an abstract object is a *relation* rather than a function or an operator.

9 *Example* To illustrate the notion of an abstract object at this point, we shall consider just one simple example involving input-output pairs which are sequences of binary symbols. To distinguish between input and output symbols, the former are denoted by 0, 1 and the latter by a, b.

The input-output pairs of lengths 1, 2, and 3 for this object, call it $\mathfrak{A}$, are listed below, with the understanding that any sequence of 0s and 1s can be an input sequence.

Input-output pairs of length 1:

$$(0,a) \quad (0,b) \quad (1,a) \quad (1,b)$$

Input-output pairs of length 2:

$(00,ab)$	$(00,ba)$	$(01,ba)$	$(01,ab)$	
$(10,aa)$	$(10,ba)$	$(11,ab)$	$(11,ba)$	$(11,bb)$

Input-output pairs of length 3:

$(000,aba)$	$(010,baa)$	$(100,bab)$	$(111,aba)$	$(000,bab)$	$(001,bab)$
$(111,bba)$	$(110,bba)$	$(111,bba)$	$(001,aba)$	$(110,baa)$	$(011,abb)$
$(100,aab)$	$(101,aab)$	$(010,aba)$	$(101,bab)$	$(110,aba)$	$(011,bab)$

We could expand this list indefinitely by listing input-output pairs of lengths 4, 5, etc. In this particular case, however, it is easy to show that the input-output pairs of length 3 determine all input-output pairs of longer length. An abstract object with this property (i.e., one in which input-output pairs of finite length k determine all input-output pairs of longer length) will be said to have *finite memory*. More concretely, a finite memory object will be said to have memory of length k if given any input-output pair $(\mathbf{u}^0_{[t_0,t_1]},\mathbf{y}^0_{[t_0,t_1]})$ of length k (i.e., one in which $t_1 - t_0 = k$) and given any continuation $\mathbf{u}^1_{(t_1,t_2]}$ of $\mathbf{u}^0_{[t_0,t_1]}$ in $\mathfrak{D}(\mathfrak{A})$ [i.e., an input $\mathbf{u}^1$ such that $\mathbf{u}^0\mathbf{u}^1 \in \mathfrak{D}(\mathfrak{A})$], there is just one continuation $\mathbf{y}^1$ of $\mathbf{y}^0$ in $\mathfrak{R}(\mathfrak{A})$. In other words, if $(\mathbf{u}^0\mathbf{u}^1,\mathbf{y}^0\mathbf{y}^1) \in \mathfrak{A}$ and $(\mathbf{u}^0,\mathbf{y}^0)$ is of length k, then $\mathbf{y}^1$ is uniquely determined by $\mathbf{u}^0$, $\mathbf{y}^0$ and $\mathbf{u}^1$. Thus for a finite-memory system $\mathbf{y}^1$ can be expressed as a function of $\mathbf{u}^0$, $\mathbf{y}^0$ and $\mathbf{u}^1$

$$\mathbf{y}^1 = \mathbf{f}(\mathbf{u}^0,\mathbf{y}^0,\mathbf{u}^1)$$

and, more particularly, for $t = \ldots, -1, 0, 1, 2, \ldots$, $\mathbf{u}^1 = (t,\mathbf{u}_t)$ and $\mathbf{y}^1 = (t,\mathbf{y}_t)$, we can write

$$\mathbf{y}_t = f(\mathbf{u}_{t-k}, \ldots ,\mathbf{u}_{t-1},\mathbf{y}_{t-k}, \ldots ,\mathbf{y}_{t-1},\mathbf{u}_t,t)$$

where $\mathbf{u}_t$ and $\mathbf{y}_t$ denote the values of $\mathbf{u}$ and $\mathbf{y}$ at time t.

To show that the abstract object under consideration has finite memory, it is sufficient to invoke the CUS condition. Specifically, consider an input-output pair of length 3 such as $(010,aba)$. From this sequence, we can deduce an input-output pair of length 4 of the form $(010u^1,abay^1)$ as follows. For the input sequence 0100, the output sequence can be either *abaa* or *abab*. Now, by the CUS condition, the last three symbols in this pair must be an input-output pair. In the list of input-output pairs, we find $(100,bab)$ but not $(100,baa)$. Consequently, $(0100,abab)$ belongs to $\mathfrak{A}$, while $(0100,abaa)$ does not. In a similar way, all input-output pairs of length greater than 3 can be deduced from pairs of length 3.

In the foregoing, we have indicated how the CUS condition can be employed to deduce input-output pairs of length >3 from those of length 3. Clearly, all pairs of shorter length can also be deduced from those of length 3 by a trivial application of the CUS condition. It is easy to verify that all input-output pairs of length 1 and 2 in the given list are segments of input-output pairs of length 3.

Input-output relations

In general, it is not practicable to list all input-output pairs which define an abstract object $\mathcal{A}$ since such pairs are usually not finite in number. For this reason, it is customary to characterize $\mathcal{A}$ by an *input-output relation,* that is, by an equation or an algorithm which can be used to generate all input-output pairs belonging to $\mathcal{A}$.

The most common example of an input-output relation is a differential equation. Thus, when we say that $\mathcal{A}$ is characterized by the input-output relation

10 $$a_n(t)\,\frac{d^ny}{dt^n} + \cdots + a_0(t)y = b_n(t)\,\frac{d^nu}{dt^n} + \cdots + b_0(t)u$$

which is a linear differential equation in u and y, we mean that any ordered pair (u,y) which satisfies *10* is an input-output pair for $\mathcal{A}$, and conversely any $(u,y) \in \mathcal{A}$ satisfies *10*.†

To illustrate, suppose that the input-output relation for $\mathcal{A}$ is expressed by

11 $$\frac{dy}{dt} + y = u$$

The general solution of this equation yields an explicit expression for all input-output pairs which satisfy *11*. Specifically, such input-output pairs can be expressed as $(u(t),y(t))$, $t \geq t_0$, in which $y(t)$ is related to u by

12 $$y(t) = \alpha e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)}u(\xi)\,d\xi$$

where α can be any real number. [Note that from setting $t = t_0$ in *12* it follows that $\alpha = y(t_0)$.] In this representation, to each value of α (which will subsequently be called a state of $\mathcal{A}$) corresponds a set of input-output pairs defined by *12*. We will make use of this observation in our later discussion of the notions of aggregate and state.

Another common example of an input-output relation is a difference equation relating input and output sequences. For example, the difference equation

13 $$y_{t+1} - cy_t = u_t \qquad t = \ldots, -1, 0, 1, \ldots$$

† Although we have not stated so explicitly, it is understood that $u(t)$ and $y(t)$ range over R^1 and that u and y are time functions on which the operations involved in *10* can be performed.

is an input-output relation in the sense that it defines all input-output pairs belonging to an object $\mathcal{A}$. More explicitly, such input-output pairs are expressed by (u_t,y_t), $t \geq t_0$, where y_t is related to u by the solution of *13*, namely

14
$$y_t = \alpha c^{t-t_0} + \sum_{\xi=t_0}^{t} c^{t-\xi-1} u_\xi \qquad t \geq t_0$$

in which α $[= y(t_0)]$ ranges over R^1. As in *11*, to each value of α (which will subsequently be called a state of $\mathcal{A}$) corresponds a set of input-output pairs of $\mathcal{A}$ defined by *14*.

Still another way of defining sets of input-output sequences is provided by graphs such as shown in Fig. *1.2.2*. In this case, u_t and y_t

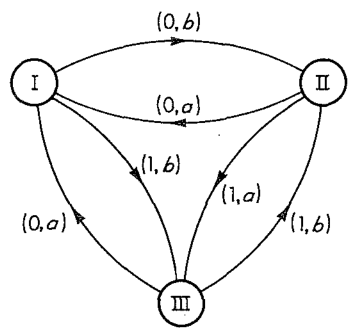


Fig. 1.2.2 State diagram for a finite-state system.

range over finite sets which for simplicity are taken to be $\{0,1\}$ and $\{a,b\}$, respectively, and each transition from one node (represented by a circle) to another is associated with an input-output pair of length 1. For example, starting in node I and applying the input sequence 01101, one gets the output sequence *babab*. This generates the input-output pair (01101,*babab*). All other input-output pairs which belong to the object in question can be generated in a similar fashion.

Objects whose input-output relations can be characterized in this fashion are called *finite-state systems.* A graph such as shown in Fig. *1.2.2* is called a *state diagram* and, as will be seen later, the nodes of such a graph play the role of the states of the object which it defines.

It should be noted that most of the abstract objects which we used for illustrative purposes in this section fall into the category of time-invariant linear systems. Since the notions of time invariance and linearity play important roles in system theory, it will be helpful at this juncture to relate these notions to the conception of an abstract object as a collection of input-output pairs. However, since we are not concerned in this chapter with the properties of any particular class of systems, our discussion of time invariance and linearity will be very brief and limited essentially to definitions of these terms. More detailed discussions of the implications of time invariance and linearity may be found in [1] and other texts and papers dealing with linear and nonlinear system theory [13–26].

Time invariance

As a preliminary to defining time invariance, it is convenient to introduce a notation for translates of a segment. Specifically, let $\mathbf{v} = \mathbf{v}_{[t_0,t_1]}$

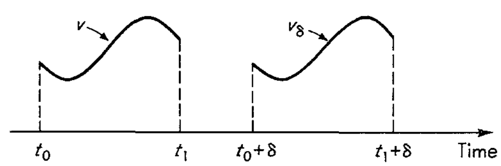


Fig. 1.2.3 Translate of a segment.

be a segment. Then, by a *translate of* $\mathbf{v}$ *by amount* δ is meant a segment $\mathbf{v}_\delta$ such that (see Fig. *1.2.3*)

15 $$\mathbf{v}_\delta(t) \triangleq \mathbf{v}(t - \delta) \qquad t_0 + \delta \leq t \leq t_1 + \delta$$

Roughly speaking, time invariance of an abstract object has to do with invariance of its input-output pairs under translations in time. More specifically, let $(\mathbf{u},\mathbf{y})$ be an input-output pair for an abstract object $\mathcal{A}$ and let $(\mathbf{u}_\delta,\mathbf{y}_\delta)$ denote a translate of $(\mathbf{u},\mathbf{y})$ by amount δ. Then $\mathcal{A}$ is a *time-invariant* abstract object if and only if the following holds true

16 $$(\mathbf{u},\mathbf{y}) \in \mathcal{A} \Rightarrow (\mathbf{u}_\delta,\mathbf{y}_\delta) \in \mathcal{A} \qquad \text{for all real } \delta$$

In other words, $\mathcal{A}$ is time-invariant if and only if the set $\mathcal{A} = \{(\mathbf{u},\mathbf{y})\}$ is closed under all translations in time.

It is easy to verify that the abstract objects defined by the input-output relations *11* and *13* are time-invariant. More generally, any input-output relation which has the form of a differential or difference equation with constant coefficients defines a time-invariant abstract object. It should be noted that the abstract object of Example *9* is tacitly assumed to be time-invariant because the times of occurrence of the first symbols in each input-output pair are not specified. This implies that not only the input-output pairs in Example *9* but also all their translates belong to the object under consideration.

Linearity

The importance of the notion of linearity stems from two facts: (1) Linearity, when it is present, greatly simplifies the analysis of system behavior; (2) many systems encountered in the real world are linear, at least to a first approximation.

Linearity can be defined quite simply in terms of the collection of input-output pairs which characterize an abstract object. Specifically, let $\mathcal{A} = \{(\mathbf{u},\mathbf{y})\}$, and let $(\mathbf{u}^0,\mathbf{y}^0)$ and $(\mathbf{u}^1,\mathbf{y}^1)$ be any pair of input-output pairs in $\mathcal{A}$. In terms of these input-output pairs, the definition of linearity can be worded as follows.

17 **Definition** An abstract object $\mathcal{A}$ is *linear* if and only if

18 $$(\mathbf{u}^0,\mathbf{y}^0) \in \mathcal{A} \quad \text{and} \quad (\mathbf{u}^1,\mathbf{y}^1) \in \mathcal{A} \Rightarrow (k_0\mathbf{u}^0 + k_1\mathbf{u}^1, k_0\mathbf{y}^0 + k_1\mathbf{y}^1) \in \mathcal{A}$$

for all real k_0 and k_1. In other words, $\mathcal{A}$ is linear if and only if any linear combination of any two input-output pairs in $\mathcal{A}$ is also an input-output pair in $\mathcal{A}$. Equivalently, $\mathcal{A}$ is linear if and only if $\mathcal{A}$ is a linear vector space.

Clearly, the input-output relations *11* and *13* define abstract objects which are linear in the sense defined above. Thus, in the case of *11*, we can write

19 $$\frac{dy^0}{dt} + y^0 = u^0$$

20 $$\frac{dy^1}{dt} + y^1 = u^1$$

and on forming a linear combination of *19* and *20*, we have

21 $$\frac{d}{dt}(k_0y^0 + k_1y^1) + k_0y^0 + k_1y^1 = k_0u^0 + k_1u^1$$

which satisfies *18* provided the domain and range of $\mathcal{A}$ are linear vector spaces, as they are tacitly assumed to be. The same applies to *13*.

3 *Some basic notions stemming from the concept of an abstract object*

This section is devoted to the discussion of several basic notions which stem from the concept of an abstract object as a collection of input-output pairs. The first of these notions is that of

1 **Containment** An object $\mathcal{A}$ will be said to be *contained* in object $\mathcal{B}$, written as $\mathcal{A} \subset \mathcal{B}$, if the set $\mathcal{A}$ is a subset of $\mathcal{B}$. In terms of input-output pairs, this means that every input-output pair belonging to $\mathcal{A}$ also belongs to $\mathcal{B}$.

2 *Example* Consider two objects $\mathcal{A}$ and $\mathcal{B}$ which are characterized by the state diagrams shown in Fig. *1.3.1*. It is easy to verify by inspection that every input-output pair which belongs to $\mathcal{A}$ also belongs to $\mathcal{B}$.

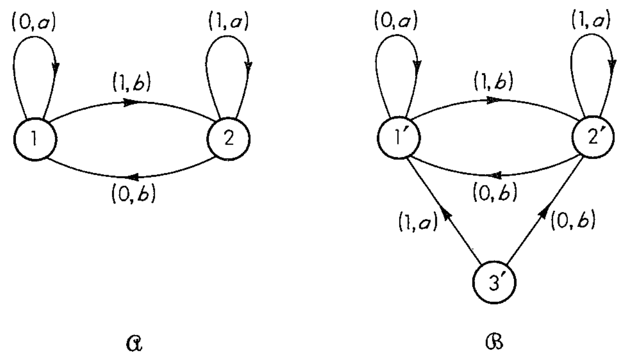


Fig. 1.3.1 State diagrams for $\mathcal{A}$ and $\mathcal{B}$, with $\mathcal{A} \subset \mathcal{B}$.

The converse, however, is not true, e.g., the input-output pair $(01,ba)$ belongs to $\mathcal{B}$ but not to $\mathcal{A}$.

3 *Example* Suppose that $\mathcal{A}$ and $\mathcal{B}$ are characterized, respectively, by the input-output relations

4 $\mathcal{A}$: $$\frac{dy}{at} = u$$

5 $\mathcal{B}$: $$\frac{d^2y}{dt^2} = \frac{du}{dt}$$

In this case, the input-output pairs for $\mathcal{A}$ and $\mathcal{B}$ may be expressed in the explicit form

6 $\mathcal{A}$: $$\left(u(t),\, \alpha + \int_{t_0}^{t} u(\xi)\, d\xi\right) \qquad t \geq t_0$$

7 $\mathcal{B}$: $$\left(u(t),\, \beta_0 + \beta_1(t - t_0) + \int_{t_0}^{t} u(\xi)\, d\xi\right) \qquad t \geq t_0$$

where α, β_0, and β_1 range over R^1. Clearly, every input-output pair belonging to $\mathcal{A}$ also belongs to $\mathcal{B}$. Hence, $\mathcal{A} \subset \mathcal{B}$.

The notion of containment leads to a very basic concept, namely, the concept of *equivalence*.[1]

[1] The concept of equivalence as defined here corresponds to that of weak equivalence in [1]. What is referred to therein as *equivalence* will be called *strong equivalence* in this chapter.

8 **Definition** $\mathcal{A}$ and $\mathcal{B}$ are equivalent, written as $\mathcal{A} = \mathcal{B}$, if $\mathcal{A}$ and $\mathcal{B}$ are equal as sets, that is, if $\mathcal{A} \subset \mathcal{B}$ and $\mathcal{B} \subset \mathcal{A}$. In words, this means that $\mathcal{A}$ and $\mathcal{B}$ are equivalent if every input-output pair which belongs to $\mathcal{A}$ also belongs to $\mathcal{B}$ and vice versa.

9 *Example* Let $\mathcal{A}$ be characterized by the system of differential equations

10 $$\mathcal{A}: \qquad \frac{dy}{dt} = v \qquad \frac{dv}{dt} = u$$

and let $\mathcal{B}$ be characterized by the differential equation

11 $$\mathcal{B}: \qquad \frac{d^2y}{dt^2} = u$$

It is easy to verify that the input-output pairs for $\mathcal{A}$ and $\mathcal{B}$ are expressed, respectively, by

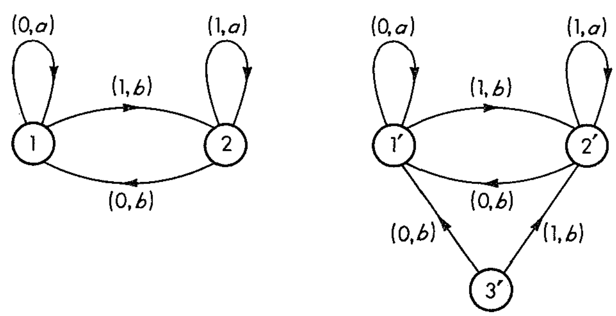
12 $$\left(u(t),\ \alpha_0 + \alpha_1(t - t_0) + \int_{t_0}^{t} d\lambda \int_{t_0}^{\lambda} u(\xi)\, d\xi\right) \qquad t \geq t_0$$

where α_0 and α_1 range over R^1, and

13 $$\left(u(t),\ \beta_0 + \beta_1(t - t_0) + \int_{t_0}^{t} d\lambda \int_{t_0}^{\lambda} u(\xi)\, d\xi\right) \qquad t \geq t_0$$

where β_0 and β_1 range over R^1. Clearly, every input-output pair which belongs to $\mathcal{A}$ also belongs to $\mathcal{B}$ and vice versa. Hence, $\mathcal{A} = \mathcal{B}$.

14 *Example* Let $\mathcal{A}$ and $\mathcal{B}$ be the objects characterized by the state diagrams shown in Fig. *1.3.2*. It is easy to verify that every input-output pair which belongs to $\mathcal{A}$ also belongs to $\mathcal{B}$ and vice versa. Hence, $\mathcal{A} = \mathcal{B}$.



$\mathcal{A}$ $\mathcal{B}$

Fig. 1.3.2 Example of equivalent systems.

Equivalence and indistinguishability

Consider the following identification problem. Suppose that an experimenter is given a black box which is known to contain either $\mathcal{A}$ or $\mathcal{B}$, with the input-output relations of $\mathcal{A}$ and $\mathcal{B}$ known to the experimenter. The experimenter is free to apply to the black box any input $\mathbf{u}$ which is admissible for both $\mathcal{A}$ and $\mathcal{B}$ (i.e., a $\mathbf{u}$ such that $\mathbf{u} \in \mathfrak{D}(\mathcal{A}) \cap \mathfrak{D}(\mathcal{B})$ and observe the response $\mathbf{y}$. Then, based on the knowledge of the input-output pair $(\mathbf{u},\mathbf{y})$, the experimenter is supposed to decide whether the black box contains $\mathcal{A}$ or $\mathcal{B}$. If the experimenter can not determine this from $(\mathbf{u},\mathbf{y})$, no matter what $\mathbf{u}$ is chosen and what $\mathbf{y}$ is observed, then $\mathcal{A}$ and $\mathcal{B}$ are said to be *indistinguishable under a single experiment.*

From the definition of equivalence, it follows at once that equivalence implies and is implied by indistinguishability under a single experiment. However, it is possible that two objects are indistinguishable under a single experiment and yet are distinguishable under a multiple experiment, that is, an experiment in which the experimenter has at his disposal more than one copy of the black box. For example, the objects shown in Fig. *1.3.2* are equivalent and yet they are distinguishable under a multiple experiment in the sense that if the experimenter has two identical copies of the black box containing, say, $\mathcal{B}$, then starting in node 3′ one copy would produce b in response to 0 while the other copy would produce b in response to 1. These responses could not be obtained if the same inputs were applied to two identical copies of $\mathcal{A}$. This implies that $\mathcal{A}$ and $\mathcal{B}$ are distinguishable by the multiple experiment in question even though they are indistinguishable under a single experiment.

As will be seen later, indistinguishability under a multiple experiment corresponds to what will be called *strong equivalence,* denoted by $\mathcal{A} \equiv \mathcal{B}$. Strong equivalence implies equivalence, but not vice versa.

4 *The notion of a system*

So far, our discussion has been concerned with a single abstract object $\mathcal{A}$ associated with input $\mathbf{u}$ and output $\mathbf{y}$. To be more general, we have to consider collections of abstract objects $\mathcal{A}_1, \ldots, \mathcal{A}_N$ in which some of the inputs or outputs associated with, say, $\mathcal{A}_i$, may be constrained to be equal (for all t) to some of the inputs or outputs of other objects in the collection. Such a combination of abstract objects will be called a *system.* It should be noted that, under this definition, every abstract object is a system and every system is an abstract object. For this reason, we shall henceforth use the term *system* to describe both an abstract object and a collection of abstract objects.

A simple example of a combination of abstract objects which form a system is shown in Fig. *1.4.1*. Here, the system $\mathcal{A}$ (represented by

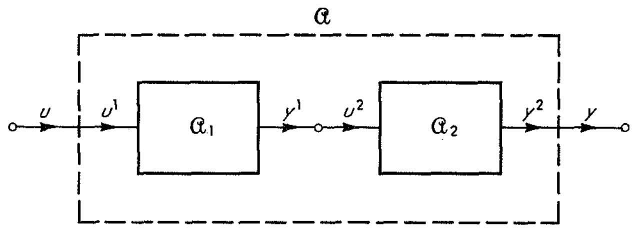


Fig. 1.4.1 Tandem combination of $\mathcal{A}_1$ and $\mathcal{A}_2$.

dotted block) with input u and output y comprises two objects, $\mathcal{A}_1$ and $\mathcal{A}_2$, connected as shown in the diagram. This connection represents the following constraint between the inputs and outputs of $\mathcal{A}_1$ and $\mathcal{A}_2$: $y^1 = u^2$, with the input and output of $\mathcal{A}$ related to those of $\mathcal{A}_1$ and $\mathcal{A}_2$ by $u = u^1$ and $y = y^2$.

A question which naturally arises when one deals with a combination of two or more systems is the following: Suppose that a system $\mathcal{A}$ is specified as a given combination of N component systems, $\mathcal{A}_1, \ldots, \mathcal{A}_N$, each of which is defined as a specified set of input-output pairs. How can one deduce from the knowledge of these sets of input-output pairs and the way in which the components of $\mathcal{A}$ (that is, $\mathcal{A}_1, \ldots, \mathcal{A}_N$) are combined the set of input-output pairs which constitutes $\mathcal{A}$? This question presents one of the central problems of system theory. We shall refer to it as the problem of *input-output analysis*.

There are two basic ways in which the problem of input-output analysis can be formulated. Essentially, the difference between these formulations lies in the way in which the sets of input-output pairs constituting the components of the given system are defined. Specifically, in the *explicit form*, these sets are defined by their respective characteristic functions, i.e., by functions which assign the value 1 to each input-output pair which belongs to a particular component system and 0 to those that do not. More concretely, if a generic input-output pair associated with a component system $\mathcal{A}_i$, $i = 1, \ldots, N$, is denoted by (u^i,y^i), then $\mathcal{A}_i$ is defined explicitly by the characteristic function $\mu_i(u^i,y^i)$:

1 $$\begin{aligned} \mu_i(u^i,y^i) &= 1 \quad \text{for } (u^i,y^i) \in \mathcal{A}_i \\ &= 0 \quad \text{for all ordered pairs of time functions which do not belong to } \mathcal{A}_i \end{aligned}$$

Thus, the knowledge of the characteristic function of $\mathfrak{A}_i$ is equivalent to having a listing or explicit characterization of all input-output pairs which belong to $\mathfrak{A}_i$.

In the *implicit form*, the component systems are defined by their respective input-output relations rather than by characteristic functions. In this case, in order to determine the input-output pairs belonging to, say, $\mathfrak{A}_i$, it is necessary to "solve" the input-output relation defining $\mathfrak{A}_i$. The problem of input-output analysis is labeled *implicit* in this case because the characterization of $\mathfrak{A}_i$ by an input-output relation is implicit rather than explicit in nature.

It will be helpful at this juncture to summarize the two formulations of the problem of input-output analysis in parallel terms.

2 **Problem of input-output analysis in explicit form** Given a system $\mathfrak{A}$ as a specified combination of component systems $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$, with each $\mathfrak{A}_i$, $i = 1, \ldots, N$, defined by a characteristic function $\mu_i(u^i,y^i)$. Determine the characteristic function of $\mathfrak{A}$ from (1) the knowledge of the characteristic functions of $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$, and (2) the constraints on the inputs and outputs of $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$ imposed by the way in which they are combined.

3 **Problem of input-output analysis in implicit form** Given a system $\mathfrak{A}$ as a specified combination of component systems $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$, with each $\mathfrak{A}_i$, $i = 1, \ldots, N$, defined by an input-output relation (e.g., a differential equation). Determine the corresponding input-output relation for $\mathfrak{A}$ from (1) the knowledge of the input-output relations for $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$, and (2) the constraints on the inputs and outputs of $\mathfrak{A}_1, \ldots, \mathfrak{A}_N$ imposed by the way in which they are combined.

As we shall see presently, on a purely formal—but not necessarily computational—level, problem *2* is much easier to solve than problem *3*. On an analytical level, problem *3* can be solved completely only for certain types of systems, e.g., systems defined by differential equations with constant coefficients.

To illustrate this point, consider a tandem combination $\mathfrak{A}$ of two systems $\mathfrak{A}_1$ and $\mathfrak{A}_2$ (Fig. *1.4.1*), in which $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are defined by their respective characteristic functions

4
$$\begin{aligned} \mu_1(u^1,y^1) &= 1 \quad \text{if } (u^1,y^1) \in \mathfrak{A}_1 \\ &= 0 \quad \text{if } (u^1,y^1) \notin \mathfrak{A}_1 \end{aligned}$$

5
$$\begin{aligned} \mu_2(u^2,y^2) &= 1 \quad \text{if } (u^2,y^2) \in \mathfrak{A}_2 \\ &= 0 \quad \text{if } (u^2,y^2) \notin \mathfrak{A}_2 \end{aligned}$$

Here the interconnection constraint is expressed by $u^2 = y^1$, with the input and output of $\mathfrak{A}$ identified with u^1 and y^2, respectively.

The constraint $u^2 = y^1$ implies that (u^1,y^2) is an input-output pair for $\mathcal{A}$ if and only if there exists a time function y^1 such that (u^1,y^1) and (y^1,y^2) are input-output pairs for $\mathcal{A}_1$ and $\mathcal{A}_2$, respectively. To express this in a more compact form, we note that the statement "$(u^1,y^1) \in \mathcal{A}_1$ and $(y^1,y^2) \in \mathcal{A}_2$" is expressible as the equation[1]

6 $$\min\,[\mu_1(u^1,y^1),\mu_2(y^1,y^2)] = 1$$

where min $[a,b]$ denotes the smaller of the two numbers a, b, with the understanding that min $[a,a] = a$. Furthermore, the statement "There exists y^1 such that $(u^1,y^1) \in \mathcal{A}_1$ and $(y^1,y^2) \in \mathcal{A}_2$" can be expressed as

7 $$\max_{y^1}\,\min\,[\mu_1(u^1,y^1),\mu_2(y^1,y^2)] = 1$$

This implies that the characteristic function of $\mathcal{A}$ can be expressed in terms of those of $\mathcal{A}_1$ and $\mathcal{A}_2$ as follows

8 $$\mu(u^1,y^2) = \max_{y^1}\,\min\,[\mu_1(u^1,y^1),\mu_2(y^1,y^2)]$$

The above relation defines the set of input-output pairs of the tandem combination of $\mathcal{A}_1$ and $\mathcal{A}_2$ in terms of the sets of input-output pairs of $\mathcal{A}_1$ and $\mathcal{A}_2$. Thus, in principle, *8* provides a solution to the problem of input-output analysis for the case where $\mathcal{A}_1$ and $\mathcal{A}_2$ are connected in tandem and the sets of input-output pairs for $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}$ are described by their respective characteristic functions.

The same approach to the problem of input-output analysis can be used in the more general case where a system $\mathcal{A}$ is an arbitrary combination of a finite number of objects $\mathcal{A}_1, \ldots, \mathcal{A}_N$. To illustrate, consider the system shown in Fig. *1.4.2*, which is composed of three interconnected objects $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$. In this case, the input u and output y of $\mathcal{A}$ are identified with u_2^1 and y^3, respectively, and the interconnection constraints read

$$y_1^1 = u_1^2 \qquad y_2^1 = u_1^3 \qquad y_1^3 = u_2^2 \qquad y_1^2 = u_1^1 \qquad y_2^2 = u_2^3$$

Now let $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$ be defined, respectively, by their characteristic functions $\mu_1(u_1^1,u_2^1;y_1^1,y_2^1)$, $\mu_2(u_1^2,u_2^2;y_1^2,y_2^2)$, $\mu_3(u_1^3,u_2^3;y_1^3)$.[2]

[1] Alternatively and more simply, we could write $\mu_1(u^1,y^1)\mu_2(y^1,y^2) = 1$. We use min rather than the product because the former has wider generality (e.g., is applicable when, as in the case of fuzzy systems (see [24]), the characteristic functions take values in the interval $[0,1]$). Note that for binary variables min $[a,b] = ab$ and max $[a,b] = a + b - ab$, where $+$ denotes sum. More generally, one can employ the notation min $[a,b] = a \wedge b$ and max $[a,b] = a \vee b$ to simplify the writing of equations such as *8* and *9*.

[2] When a system has multiple inputs and/or outputs, it is helpful to use a semicolon in the characteristic function to separate the input variables from the output variables.

Then, by inspection, the characteristic function for $\mathfrak{A}$ can be expressed in terms of μ_1, μ_2, and μ_3 as follows:

9 $$\mu(u_2^1,y_1^3) = \max_{y_1^1,\, y_2^1,\, y_1^2,\, y_2^2} \min\,[\mu_1(y_1^2,u_2^1;y_1^1,y_2^1),\mu_2(y_1^1,y_1^3;y_1^2,y_2^2),\mu_3(y_2^1,y_2^2;y_1^3)]$$

As was pointed out earlier, this equation is merely a compact expression for the statement "An input-output pair (u_2^1,y_1^3) belongs to $\mathfrak{A}$ if and only if there exist y_1^1, y_2^1, y_1^2, and y_2^2 such that $(y_1^2,u_2^1;y_1^1,y_2^1)$, $(y_1^1,y_1^3;y_1^2,y_2^2)$, and $(y_2^1,y_2^2;y_1^3)$ are input-output pairs for $\mathfrak{A}_1$, $\mathfrak{A}_2$, and $\mathfrak{A}_3$, respectively."

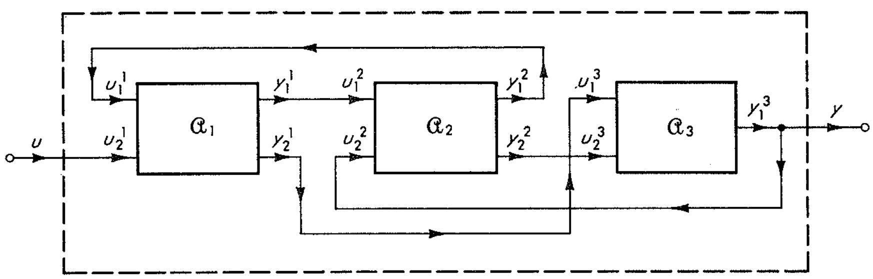


Fig. 1.4.2 Example of a system which is a combination of three component systems $\mathfrak{A}_1$, $\mathfrak{A}_2$, and $\mathfrak{A}_3$.

In the foregoing examples, we assumed that the objects comprising a system are defined by their respective characteristic functions and that the problem is to express the characteristic function of the system in terms of the characteristic functions of its components. The examples show that, in general, the problem of input-output analysis in explicit form is easy to solve, at least on a purely formal level.

Next, let us turn our attention to the case where a system is composed of objects which are defined by their respective input-output relations and the problem is to find an expression for the input-output relation for the system in terms of those of its components. For concreteness, assume that $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are connected in tandem as in Fig. *1.4.1* and that their respective input-output relations read

10 $\mathfrak{A}_1$: $$(p^2 + 3p + 2)y^1 = u^1$$

11 $\mathfrak{A}_2$: $$(p + 3)y^2 = pu^2$$

where p denotes the derivative operator d/dt. Then, as is shown in [1], the input-output relation for $\mathfrak{A}$ is given by

$$(p + 3)(p^2 + 3p + 2)y^2 = pu^1$$

More generally, if the input-output relations for $\mathcal{A}_1$ and $\mathcal{A}_2$ are of the form

12 $$\mathcal{A}_1: \qquad L_1(p)y^1 = M_1(p)u^1$$

13 $$\mathcal{A}_2: \qquad L_2(p)y^2 = M_2(p)u^2$$

where L_1, M_1, L_2, and M_2 are differential operators such that L_1 and M_2 have no factors in common, then it can be proved [1] that the input-output relation for $\mathcal{A}$ is given by

14 $$L_1(p)L_2(p)y^2 = M_1(p)M_2(p)u^1$$

As will be shown in Chap. *5*, the problem of input-output analysis in implicit form can be solved completely when $\mathcal{A}_1, \ldots, \mathcal{A}_N$ are systems whose input-output relations are linear differential equations with constant coefficients. Unfortunately, this is just about the only case in which the problem of input-output analysis in implicit form can be solved analytically. In particular, no analytic procedure is available for the solution of this problem when the input-output relations of $\mathcal{A}_1, \ldots, \mathcal{A}_N$ are linear differential equations with time-varying coefficients. For example, in the case of the tandem combination of Fig. *1.4.1*, suppose that $\mathcal{A}_1$ and $\mathcal{A}_2$ are characterized by input-output relations of the form

15 $$\mathcal{A}_1: \quad [a_n^1(t)p^n + \cdots + a_0^1(t)]y^1 = [b_n^1(t)p^n + \cdots + b_0^1(t)]u^1$$

16 $$\mathcal{A}_2: \quad [a_n^2(t)p^n + \cdots + a_0^2(t)]y^2 = [b_n^2(t)p^n + \cdots + b_0^2(t)]u^2$$

where the $a_i^1(t)$, $a_i^2(t)$, $b_i^1(t)$, $b_i^2(t)$, $i = 1, \ldots, n$, are specified time-varying coefficients. In this case, there is no general analytic procedure for expressing the input-output relation for $\mathcal{A}$ in the form of a differential equation

17 $$L(p,t)y^2 = M(p,t)u^1$$

where $L(p,t)$ and $M(p,t)$ are polynomials in p with time-varying coefficients.

This concludes our brief discussion of the notion of a system and the problem of input-output analysis. We are now ready to proceed to the definition of the notions of aggregate and state.

5 *Aggregates and states*

Although the notion of state has a long history of use in many fields of science, especially in analytic dynamics, thermodynamics, and quantum mechanics, it is hard to find a definition of it in the scientific literature which does not suffer either from narrowness or imprecision. In what follows, we shall show how the concept of state can be given a precise and yet very general meaning by interpreting it as a tag attached to a subset of input-output pairs in $\mathcal{A}$. The reader must be warned, however, that the naturalness of this point of view will not be apparent in the initial stages of our discussion.

Let $\mathcal{A}$ be a system defined as a set of input-output pairs

$$\mathcal{A} = \{(u,y)\}$$

satisfying the CUS condition (see Sec. *2*). For each t_0, let $\mathcal{A}(t_0)$ denote the subset of $\mathcal{A}$ comprising all input-output pairs which start at t_0, that is,

1 $$\mathcal{A}(t_0) = \{(\mathbf{u}_{[t_0,t]}, \mathbf{y}_{[t_0,t]})\} \qquad t \geq t_0$$

It would be natural to attempt to group together those input-output pairs in $\mathcal{A}(t_0)$ which exhibit a particular characteristic or have some specified property in common. With this in view, let $\mathcal{A}_{\alpha_0}(t_0)$ denote a specified subset of input-output pairs in $\mathcal{A}(t_0)$, with the index $\boldsymbol{\alpha}_0$ serving as an identifying label. For suggestiveness, such a subset will be referred to as a *bundle of input-output pairs* or, simply, a *bundle*, and $\boldsymbol{\alpha}_0$ will be called its *tag*.

To illustrate, consider a system $\mathcal{A}$ defined by the input-output relation

$$\frac{dy}{dt} + y = u$$

By solving this equation for y, we can express $\mathcal{A}(t_0)$ explicitly as

$$\mathcal{A}(t_0) = \{\mathcal{A}_{\alpha_0}(t_0)\}_{\alpha_0} \qquad \alpha_0 \in (-\infty, \infty)$$

where

2 $$\mathcal{A}_{\alpha_0}(t_0) = \left\{u(t),\ \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)} u(\xi)\, d\xi,\ t \geq t_0\right\}_u$$

and $u(t) \in (-\infty, \infty)$. In this case, for each α_0, *2* defines a bundle of input-output pairs in $\mathcal{A}(t_0)$, with u being the generating variable.

As in the above example, usually our concern will be not with a single bundle $\mathcal{A}_{\alpha_0}(t_0)$, but with a family of bundles, $\{\mathcal{A}_{\alpha_0}(t_0)\}$, for each t_0. Such a family would be generated by permitting $\boldsymbol{\alpha}_0$ to vary over some space Σ_{t_0}. The subscript t_0 in Σ_{t_0} serves to anticipate the possibility that the range of $\boldsymbol{\alpha}_0$ may be dependent on t_0. When Σ_{t_0} is independent of t_0, it will be denoted by Σ.[1]

[1] In terms of the notation described in Notational Preliminaries (Sec. *1*), $\{\mathcal{A}_{\alpha_0}(t_0)\}$ should be interpreted as $\{\mathcal{A}_{\alpha_0}(t_0)\}_{\alpha_0}$, that is, a family of bundles of input-output pairs starting at t_0, with $\boldsymbol{\alpha}_0$ being the generating variable. When both t_0 and $\boldsymbol{\alpha}_0$ are varied, the resulting family of bundles will be denoted by $\{\mathcal{A}_{\alpha_0}(t_0)\}_{\alpha_0,t_0}$. A bundle in this family which comprises input-output pairs starting at, say, t_1, will be denoted by $\mathcal{A}_{\alpha_1}(t_1)$, where $\boldsymbol{\alpha}_1$ is a tag for the bundle in question and $\boldsymbol{\alpha}_1 \in \Sigma_{t_1}$.

To recapitulate, for each t_0 in $(-\infty, \infty)$ we assume that in some as yet unspecified way the input-output pairs in $\mathcal{A}(t_0)$ are grouped (bundled) together on the basis of some property which they have in common. A generic bundle is denoted by $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$, with its identifying tag $\boldsymbol{\alpha}_0$ being an element of a specified space Σ_{t_0}. For each t_0, varying $\boldsymbol{\alpha}_0$ over Σ_{t_0} generates a family of bundles which we denote by $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$. The members of this family are subsets of input-output pairs in $\mathcal{A}(t_0)$. These subsets need not be disjoint.

So far we have not made any restrictive assumptions regarding the ways in which the input-output pairs in $\mathcal{A}(t_0)$ may be grouped together into bundles. Now, to pave the way for the introduction of the notion of state, we shall focus our attention on a special way of bundling the input-output pairs in $\mathcal{A}(t_0)$. The resulting bundles will be referred to as *aggregates of input-output pairs* or, simply, *aggregates*, and their tags will constitute the *states* of $\mathcal{A}$. As will become more apparent later, the concept of an aggregate is intrinsically more basic than that of state, mainly because a state is merely a tag for an aggregate and as such is less intrinsic in relation to $\mathcal{A}$.

In order to qualify to be called aggregates, the bundles in the family $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$, $\boldsymbol{\alpha}_0 \in \Sigma_{t_0}$, must satisfy four consistency conditions which are set forth below. The motivation for these conditions will become clear once we have defined the notion of state.

3 **Covering** This condition requires that, for each t_0, the family of bundles $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}_{\boldsymbol{\alpha}_0, t_0}$ be a covering for $\mathcal{A}(t_0)$ in the sense that the subset of input-output pairs starting at t_0, $\mathcal{A}(t_0)$, be the union of the bundles $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$, $\boldsymbol{\alpha}_0 \in \Sigma_{t_0}$, that is,

4
$$\bigcup_{\boldsymbol{\alpha}_0} \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0) = \mathcal{A}(t_0)$$

The purpose of this condition is to insure that every input-output pair in $\mathcal{A}(t_0)$ belongs to at least one bundle in the family $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$, $\boldsymbol{\alpha}_0 \in \Sigma_{t_0}$.

5 **Closure under truncation** Let $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1)$ be an arbitrary input-output pair in $\mathcal{A}(t_0)$, with $(\mathbf{u}^0, \mathbf{y}^0)$ representing a left truncate of $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1)$.† Then, the condition in question requires that if $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1)$ is in a bundle $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$, so must be every left truncate of this pair. In symbols

6
$$(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1) \in \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0) \Rightarrow (\mathbf{u}^0, \mathbf{y}^0) \in \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$$

for all $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1) \in \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$, all t_0, and all $\boldsymbol{\alpha}_0 \in \Sigma_{t_0}$.

† In accordance with the notation introduced in Sec. *1*, $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1)$ is an input-output pair comprising an input-output pair $(\mathbf{u}^0, \mathbf{y}^0)$ followed by an input-output pair $(\mathbf{u}^1, \mathbf{y}^1)$. The input-output pairs $(\mathbf{u}^0, \mathbf{y}^0)$ and $(\mathbf{u}^1, \mathbf{y}^1)$ are, respectively, left and right truncates of the input-output pair $(\mathbf{u}^0\mathbf{u}^1, \mathbf{y}^0\mathbf{y}^1)$.

7 **Uniqueness** By analogy with the domain and range of a system (see Sec. *2*), we can define the domain and range of a bundle of input-output pairs $\mathcal{A}_{\alpha_0}(t_0)$ as the sets

8
$$\mathfrak{D}(\mathcal{A}_{\alpha_0}(t_0)) \triangleq \{\mathbf{u}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}_{\alpha_0}(t_0)\}$$

and

9
$$\mathcal{R}(\mathcal{A}_{\alpha_0}(t_0)) \triangleq \{\mathbf{y}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}_{\alpha_0}(t_0)\}$$

The uniqueness condition requires that to each input $\mathbf{u}$ in $\mathfrak{D}(\mathcal{A}_{\alpha_0}(t_0))$ correspond a unique $\mathbf{y}$ in $\mathcal{R}(\mathcal{A}_{\alpha_0}(t_0))$. Equivalently, this condition may be expressed compactly by the implication

10
$$\{(\mathbf{u},\mathbf{y}) \in \mathcal{A}_{\alpha_0}(t_0) \text{ and } (\mathbf{u},\mathbf{y}') \in \mathcal{A}_{\alpha_0}(t_0)\} \Rightarrow \mathbf{y} = \mathbf{y}'$$

If the uniqueness condition is satisfied, then $\mathbf{y}$ can be expressed as a function of the input $\mathbf{u}$ and $\boldsymbol{\alpha}_0$, with the latter identifying the bundle $\mathcal{A}_{\alpha_0}(t_0)$ to which $(\mathbf{u},\mathbf{y})$ belongs. We shall express this by writing

11
$$\mathbf{y} = \bar{A}(\boldsymbol{\alpha}_0;\mathbf{u})$$

where $\bar{A}$ is a function from $\Sigma_{t_0} \times \mathfrak{D}(\mathcal{A}_{\alpha_0}(t_0))$ to $\mathcal{R}(\mathcal{A}_{\alpha_0}(t_0))$. Later on, this equation will be referred to as an *input-output-state relation* for $\mathcal{A}$.

12 **Continuation** Let $(\mathbf{u}^0,\mathbf{y}^0)$ be an input-output pair in $\mathcal{A}_{\alpha_0}(t_0)$ over an interval $[t_0,t_1]$, and let $(\mathbf{u}^1,\mathbf{y}^1)$ be an input-output pair in $\mathcal{A}(t_1)$ over an interval $[t_1,t]$. As in Sec. *2*, the input-output pair $(\mathbf{u}^1,\mathbf{y}^1)$ will be said to be a *continuation* of $(\mathbf{u}^0,\mathbf{y}^0)$ in $\mathcal{A}_{\alpha_0}(t_0)$ if the input-output pair $(\mathbf{u}^0\mathbf{u}^1,\mathbf{y}^0\mathbf{y}^1)$ is in $\mathcal{A}_{\alpha_0}(t_0)$. In effect, $(\mathbf{u}^1,\mathbf{y}^1)$ is merely a right truncate of an input-output pair in $\mathcal{A}_{\alpha_0}(t_0)$.

The continuation condition requires that the set of all continuations of $(\mathbf{u}^0,\mathbf{y}^0)$ in $\mathcal{A}_{\alpha_0}(t_0)$ be a bundle of input-output pairs in the family $\{\mathcal{A}_{\alpha_0}(t_0)\}_{\alpha_0,t_0}$. If the tag of this bundle is denoted by $\boldsymbol{\alpha}_1$ (with $\boldsymbol{\alpha}_1 \in \Sigma_{t_1}$), then the condition in question can be expressed as the equality

13
$$\{(\mathbf{u}^1,\mathbf{y}^1)|(\mathbf{u}^0\mathbf{u}^1,\mathbf{y}^0\mathbf{y}^1) \in \mathcal{A}_{\alpha_0}(t_0)\} = \mathcal{A}_{\alpha_1}(t_1)$$

which should hold for all t_0 in R^1, all $\boldsymbol{\alpha}_0$ in Σ_{t_0}, all $(\mathbf{u}^0,\mathbf{y}^0)$ in $\mathcal{A}_{\alpha_0}(t_0)$, and all $t_1 \geq t_0$, $t \geq t_1$ (see Fig. *1.5.1*).

To illustrate the four consistency conditions stated above, consider the time-invariant finite-state system defined in Example *1.2.9*, and assume that for each t_0, $t_0 = \ldots, -1, 0, 1, \ldots$, the input-output pairs of length $l \leq 3$ are bundled together into three groups indexed by I, II and III, that is, $\Sigma_{t_0} = \{\mathrm{I,II,III}\}$, independent of t_0. The members of the three bundles $\mathcal{A}_{\mathrm{I}}$, $\mathcal{A}_{\mathrm{II}}$, and $\mathcal{A}_{\mathrm{III}}$ [with t_0 omitted as argument in $\mathcal{A}_{\mathrm{I}}(t_0)$, $\mathcal{A}_{\mathrm{II}}(t_0)$, and $\mathcal{A}_{\mathrm{III}}(t_0)$ because the bundles are the same for all t_0] are tabulated in Table *1.5.1*.

It is easy to verify by exhaustive testing that this bundling of input-output pairs of length ≤ 3 satisfies the four consistency conditions. Specifically, for $l \leq 3$

1. The covering condition is satisfied because the union of $\mathcal{A}_{\mathrm{I}}$, $\mathcal{A}_{\mathrm{II}}$, and $\mathcal{A}_{\mathrm{III}}$ is $\mathcal{A}$.

2. The closure under truncation condition is satisfied because the left truncates of every input-output pair in each bundle are in the same bundle. For example, the left truncates of (000,*bab*), which belongs to $\mathcal{A}_{\mathrm{I}}$, are (0,*b*) and (00,*ba*), both of which belong to $\mathcal{A}_{\mathrm{I}}$.

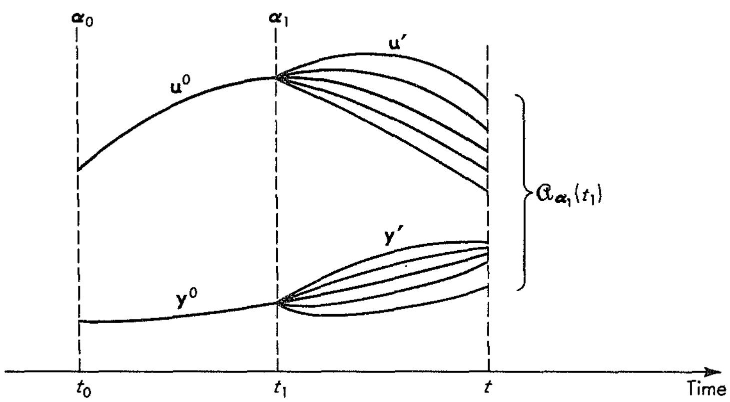


Fig. 1.5.1 Illustration of the continuation condition.

Table 1.5.1

I	II	III
(0,*b*)	(0,*a*)	(0,*a*)
(1,*b*)	(1,*a*)	(1,*b*)
(00,*ba*)	(00,*ab*)	(00,*ab*)
(01,*ba*)	(01,*ab*)	(01,*ab*)
(10,*ba*)	(10,*aa*)	(10,*ba*)
(11,*bb*)	(11,*ab*)	(11,*ba*)
(000,*bab*)	(000,*aba*)	(000,*aba*)
(001,*bab*)	(001,*aba*)	(001,*aba*)
(010,*baa*)	(010,*aba*)	(010,*aba*)
(011,*bab*)	(011,*abb*)	(011,*abb*)
(100,*bab*)	(100,*aab*)	(100,*bab*)
(101,*bab*)	(101,*aab*)	(101,*bab*)
(110,*bba*)	(110,*aba*)	(110,*baa*)
(111,*bba*)	(111,*aba*)	(111,*bab*)

3. The uniqueness condition is satisfied because, in each bundle, to every input sequence of length ≤ 3 corresponds a unique response. For example, for input-output pairs in $\mathcal{A}_{\mathrm{I}}$, to the input (000) corresponds a unique response *bab*. Likewise, for input-output pairs in $\mathcal{A}_{\mathrm{II}}$, to the input (000) corresponds a unique response (*aba*). The same is true for all input-output pairs in each bundle.

4. As an example of how the continuation condition is checked, consider an input-output pair such as $(0,b)$ in $\mathcal{A}_{\mathrm{I}}$. By inspection, the continuations (of length ≤ 2) of this input-output pair in $\mathcal{A}_{\mathrm{I}}$ are found to be:

14
$$\begin{array}{c}(0,a)\\(1,a)\\(00,ab)\\(01,ab)\\(10,aa)\\(11,ab)\end{array}$$

which are members of $\mathcal{A}_{\mathrm{II}}$. In a similar fashion, we can verify that the set of all continuations of $(0,b)$ in $\mathcal{A}_{\mathrm{I}}$ coincides with $\mathcal{A}_{\mathrm{II}}$.

An additional illustration of the four consistency conditions is provided by the following

15 *Example* Let $\mathcal{A}$ be a time-invariant system defined as a family, $\{\mathcal{A}_{\alpha_0}(t_0)\}$, of bundles of input-output pairs

16
$$\mathcal{A}_{\alpha_0}(t_0) = \{[u(t),y(t)],\ t \geq t_0\}$$

where

17
$$y(t) = \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)} u(\xi)\, d\xi$$

in which α_0 ranges over $(-\infty,\infty)$. Thus $\Sigma_{t_0} = (-\infty,\infty)$.

In this case, the covering, uniqueness, and closure under truncation conditions are satisfied by the definition of $\mathcal{A}$. To verify that the continuation condition is satisfied, let us pick an input-output pair in $\mathcal{A}_{\alpha_0}(t_0)$, say, $(u^0(t),y^0(t))$, $t_0 \leq t \leq t_1$, and demonstrate that the set of all of its continuations in $\mathcal{A}_{\alpha_0}(t_0)$ coincides with $\mathcal{A}_{\alpha_1}(t_1)$ for some α_1 in $(-\infty,\infty)$.

Specifically, let (u^1,y^1) be a continuation of (u^0,y^0) in $\mathcal{A}_{\alpha_0}(t_0)$. Then, since $(u^0u^1,y^0y^1) \in \mathcal{A}_{\alpha_0}(t_0)$, we must have

18
$$y^1(t) = \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)} u(\xi)\, d\xi \qquad t \geq t_1$$

Now, $y^1(t)$ as given by *18* can be rewritten successively as

19 $$y^1(t) = \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t_1} e^{-(t-\xi)} u^0(\xi)\, d\xi + \int_{t_1}^{t} e^{-(t-\xi)} u^1(\xi)\, d\xi \qquad t \geq t_1$$
$$= \alpha_0 e^{-(t-t_0)} + e^{-(t-t_1)} \int_{t_0}^{t_1} e^{-(t_1-\xi)} u^0(\xi)\, d\xi + \int_{t_1}^{t} e^{-(t-\xi)} u^1(\xi)\, d\xi \qquad t \geq t_1$$

20 $$= \alpha_1 e^{-(t-t_1)} + \int_{t_1}^{t} e^{-(t-\xi)} u^1(\xi)\, d\xi \qquad t \geq t_1$$

where

21 $$\alpha_1 = \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t_1} e^{-(t_1-\xi)} u^0(\xi)\, d\xi$$

From inspection of *20*, it follows at once that the set of all continuations of (u^0,y^0) in $\mathcal{A}_{\alpha_0}(t_0)$, that is,

22 $$\{(u^1(t),y^1(t)),\ t \geq t_1 | (u^0u^1,y^0y^1) \in \mathcal{A}_{\alpha_0}(t_0)\}$$

coincides with $\mathcal{A}_{\alpha_1}(t_1)$, and hence that the continuation condition is satisfied by the input-output pairs defined by *16*.

Aggregates and states

Our main purpose in introducing the notion of a bundle of input-output pairs and in formulating a set of consistency conditions which families of such bundles must satisfy was to set the stage for defining the aggregates and states of a system in terms of its input-output pairs. We are now in a position to do this concisely and in a way which adds considerable insight to the meaning of the concept of state and its basic properties.

23 **Definition** Let a system $\mathcal{A}$ be defined as a collection of input-output pairs

24 $$\mathcal{A} = \{(\mathbf{u},\mathbf{y})\}$$

satisfying the CUS condition (see Sec. *2*). Let $\{\mathcal{A}_{\alpha_0}(t_0)\}_{\alpha_0,t_0}$ be a family of bundles of input-output pairs satisfying the covering *3*, closure under truncation *5*, uniqueness *7*, and continuation *12* conditions. Such a family will be said to be *consistent*, and a bundle $\mathcal{A}_{\alpha_0}(t_0)$ in a consistent family will be said to be an *aggregate*.

25 **Definition** Let $\mathcal{A}_{\alpha_0}(t_0)$ be an aggregate in $\mathcal{A}$ at time t_0. The identifying tag $\boldsymbol{\alpha}_0$ of $\mathcal{A}_{\alpha_0}(t_0)$ is a *state of* $\mathcal{A}$ *at time* t_0. The range of $\boldsymbol{\alpha}_0$, Σ_{t_0} is the *state space of* $\mathcal{A}$ *at time* t_0. When Σ_{t_0} is independent of t_0—as is usually the case—it is denoted by Σ and is referred to more simply as the *state space of* $\mathcal{A}$.

When necessary to place in evidence the dependence of state on time, we shall denote it as $\mathbf{x}(t)$ or $\mathbf{x}_t$, according as t ranges over $(-\infty,\infty)$

or the set of integers. Thus, it will be understood that, for each t_0, $\mathbf{x}(t_0)$ is a variable ranging over Σ_{t_0}, with $\boldsymbol{\alpha}_0$ representing a generic value of $\mathbf{x}(t_0)$. On occasion, $\boldsymbol{\alpha}_0$ or $\mathbf{x}(t_0)$ will be referred to as the *initial state* of $\mathcal{A}$ at time t_0 in order to differentiate it from a state of $\mathcal{A}$ at some subsequent time t.

26 *Comment* The definition of a state of a system $\mathcal{A}$ as a tag for an aggregate in $\mathcal{A}$ underscores the fact that the states of $\mathcal{A}$ do not have a unique identity and have much less intrinsic connection with $\mathcal{A}$ than the aggregates in $\mathcal{A}$. This, however, naturally raises the question: Are the aggregates in $\mathcal{A}$ uniquely determined by the four consistency conditions *3*, *5*, *7*, and *12*? It can be shown that this question can be answered affirmatively for certain classes of systems, e.g., systems whose input-output relations have the form of linear differential equations with constant coefficients (Theorem *3.9.1* in [1]). An example to the contrary is furnished by the two finite-state systems shown in Fig. *1.3.2*. In this case, $\mathcal{A}$ and $\mathcal{B}$ have the same set of input-output pairs but not the same family of aggregates. In the terminology of finite-state systems, $\mathcal{B}$ is not strongly connected, that is, it contains states (state *3′*) which are not reachable from other states (*1′* and *2′*). If the aggregates were required to be reachable from one another [in the sense that for each $\boldsymbol{\alpha}_0$ and each $\boldsymbol{\alpha}_1$ there exists $\mathbf{u}_{[t_0,t_1]}$ such that $\mathcal{A}_{\boldsymbol{\alpha}_1}(t_1)$ is the set of all continuations of $(\mathbf{u}_{[t_0,t_1]},\mathbf{y}_{[t_0,t_1]})$ in $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$], then the family of aggregates would be uniquely determined by the four consistency conditions.

Input-output-state relations

In consequence of the uniqueness condition, to each $\mathbf{u}$ in the domain of an aggregate $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$ corresponds a unique $\mathbf{y}$ in its range. Thus, for any input-output pair in $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$ we can write

27 $$\mathbf{y} = \bar{\mathbf{A}}(\boldsymbol{\alpha}_0;\mathbf{u})$$

or equivalently

28 $$\mathbf{y} = \bar{\mathbf{A}}(\mathbf{x}(t_0);\mathbf{u})$$

where $\bar{\mathbf{A}}$ is a function from $\Sigma_{t_0} \times \mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$ to $\mathcal{R}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$. In view of their form, these relations will be referred to as the *input-output-state relations* for $\mathcal{A}$.

It should be noted that an input-output-state relation expresses an output segment $\mathbf{y}$ as a function of initial state $\boldsymbol{\alpha}_0$ and an input segment $\mathbf{u}$. Since there is no convenient way of expressing a segment of a time function as a function of other time functions, the information provided by an input-output-state relation will usually be presented in the form of an equation

29 $$\mathbf{y}(t) = \mathbf{A}(\mathbf{x}(t_0);\mathbf{u})$$

in which $\mathbf{y}(t)$, rather than $\mathbf{y}$, is expressed as a function of $\mathbf{x}(t_0)$ and $\mathbf{u}$. Like *27* and *28*, this equation will be referred to as an input-output-state relation. It is important to observe that when $\mathcal{A}$ is defined by an input-output relation having the form of a differential or difference equation, *29* expresses the general solution of such an equation, with $\mathbf{x}(t_0)$ representing the initial conditions.

Having defined the states of $\mathcal{A}$, we are now in a position to see more clearly the motivation for the less obvious consistency conditions imposed on the bundles of input-output pairs of $\mathcal{A}$. Specifically, the uniqueness condition *7* serves to insure that if $\mathcal{A}$ is initially (at t_0) in some state $\boldsymbol{\alpha}_0$, then the response of $\mathcal{A}$ to any given input $\mathbf{u}$ in $\mathfrak{D}(\mathcal{A}_{\alpha_0}(t_0))$ is uniquely determined by that input and $\boldsymbol{\alpha}_0$ through the input-output-state relation *27*. This, of course, expresses a key property of the notion of a state of a system, viz., that the knowledge of state at time t_0 is sufficient to determine the response of the system to any input starting at time t_0. In other words, the state of a system contains all the information about its past history that is relevant to the prediction of its future behavior.

The continuation condition has the following interpretation. Suppose that $\mathcal{A}$ is initially (at t_0) in some state $\boldsymbol{\alpha}_0$ and that an input $\mathbf{u}^0$, $\mathbf{u}^0 \triangleq \mathbf{u}_{[t_0,t_1]}$ is applied to $\mathcal{A}$. Then at time t_1, $\mathcal{A}$ will be in a state $\boldsymbol{\alpha}_1$ which defines uniquely an aggregate $\mathcal{A}_{\alpha_1}(t_1)$. Thus, $\boldsymbol{\alpha}_1$ can be regarded as the initial state (at time t_1) for any input starting at t_1. We shall say that $\boldsymbol{\alpha}_0$ is *taken by* $\mathbf{u}^0$ into $\boldsymbol{\alpha}_1$.

The above interpretation contains an important conclusion, namely, that $\mathcal{A}_{\alpha_1}(t_1)$ is uniquely determined by $\boldsymbol{\alpha}_0$ and $\mathbf{u}^0$. This does not imply that $\boldsymbol{\alpha}_1$ is uniquely determined by $\boldsymbol{\alpha}_0$ and $\mathbf{u}^0$ since each bundle may be tagged in a number of different ways. However, as will be seen in Sec. *6*, all such tags are equivalent to one another. Consequently, $\boldsymbol{\alpha}_1$ is uniquely determined, to within equivalent states (see *1.6.1*), by $\boldsymbol{\alpha}_0$ and $\mathbf{u}^0$. In a general form, this will be expressed by the equation

30 $$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t_1]})$$

which will be referred to as a *state equation* for $\mathcal{A}$. In this equation, $\mathbf{F}$ is a function from $\Sigma_{t_0} \times \mathfrak{D}(\mathcal{A}_{\alpha_0}(t_0))$ to Σ_t which expresses the state at time t in terms of the state at time t_0 and the input segment over the interval $[t_0,t]$.

To illustrate the notions introduced in the foregoing discussion, it will be helpful to consider two examples which were employed previously, namely, Example *1.2.9* and Example *1.5.15*.

In Example *1.2.9*, it is easy to verify by inspection of $\mathcal{A}_{\mathrm{I}}$, $\mathcal{A}_{\mathrm{II}}$, and $\mathcal{A}_{\mathrm{III}}$ that these aggregates comprise, respectively, input-output pairs starting in nodes I, II, and III. For example, the input-output pair (010,*baa*) in $\mathcal{A}_{\mathrm{I}}$ results from applying 0 starting in node I—which yields b and leads to node II; then applying 1 starting in node II—which yields a and leads to node III; and then applying 0 starting in node III—which yields a and leads to node I. In short, the input-output pair (010,*baa*) results from applying the input sequence 010 starting in node I—which yields the output sequence *baa* and leads to node I.

Combining this observation with the fact that $\mathcal{A}_{\mathrm{I}}$, $\mathcal{A}_{\mathrm{II}}$, and $\mathcal{A}_{\mathrm{III}}$ have been verified to be aggregates of input-output pairs of $\mathcal{A}$, it follows that the tags I, II, and III, which represent the nodes of the graph defining $\mathcal{A}$ (Fig. *1.2.2*), constitute the states of $\mathcal{A}$. From the graph, it follows at once that if the state of $\mathcal{A}$ at time t is denoted by x_t, then we can write

31 $$x_{t+1} = f(x_t,u_t) \qquad t = \ldots, -1, 0, 1, \ldots$$

32 $$y_t = g(x_t,u_t)$$

where f and g are tabulated below

33

u_t \ x_t	I	II	III
0	II	I	I
1	III	III	II

34

u_t \ x_t	I	II	III
0	b	a	a
1	b	a	b

By iteration, *31* can be used to yield an expression for x_t, $t \geq t_0$, in terms of x_{t_0} and $u_{t_0}, \ldots, u_{t-1}$. Thus,

35 $$x_t = f(f(\ldots,f(f(x_{t_0},u_{t_0}),u_{t_0+1}), \ldots, u_{t-1}))$$

which is a state equation for $\mathcal{A}$ in the sense of *30*. Then, on substituting this expression into *32*, we have

36 $$y_t = g(f(\ldots,f(f(x_{t_0},u_{t_0}),u_{t_0+1}), \ldots,u_{t-1}),u_t)$$

which is an input-output-state relation for $\mathcal{A}$. Thus, in combination, equations of the form *31* and *32* completely define a finite-state system and place in evidence its states and aggregates.

It should be noted that, in the literature of finite-state systems, it is customary to define a time-invariant finite-state system as a triple of finite sets U (range of u_t), Y (range of y_t), and Σ (range of x_t), and the pair of functions f and g defined by *31* and *32*. Thus, what is the point of departure in the conventional approach is a terminal point in the approach used in this chapter.

Turning to Example *15*, we note that in this case the input-output state relation *29* reads

37 $$y(t) = x(t_0)e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)}u(\xi)\,d\xi \qquad t \geq t_0$$

This equation is obtained from the definition of $\mathcal{A}_{\alpha_0}(t_0)$ (see *16*) merely by replacing α_0 with $x(t_0)$ in *17*.

From *21*, it follows that the state equation *30* can be expressed in the explicit form

38 $$x(t) = x(t_0)e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)}u(\xi)\,d\xi \qquad t \geq t_0$$

Note that in the example under consideration $y(t) = x(t)$ for all t. More generally, the output at time t is a function of the input at time t and the state at time t, that is,

39 $$y(t) = g(x(t),u(t),t)$$

This equation results from *29* by letting $t_0 = t$ in the input-output-state relation. Conversely, the input-output-state relation can be obtained from *39* by combining *39* with the state equation *30*, that is, by substituting the right-hand member of *30* in place of $\mathbf{x}(t)$ in *39*, yielding

40 $$\mathbf{y}(t) = \mathbf{g}(\mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}),\mathbf{u}(t),t)$$

which expresses $\mathbf{y}(t)$ as a function of $\mathbf{x}(t_0)$ and $\mathbf{u}_{[t_0,t]}$.

6 *State and system equivalence*

The preceding sections were devoted in the main to the introduction of three basic concepts: System, aggregate, and state. In essence, we have defined a system as a collection of input-output pairs; we have defined an aggregate as a bundle of input-output pairs satisfying certain consistency conditions; and we have defined a state as a tag attached to an aggregate.

As was pointed out earlier, since a state is merely a tag or a name for an aggregate, it is not as intrinsic a concept as that of an aggregate, in the sense that the same aggregate can be described by a variety of different tags whereas to each tag will correspond but one aggregate. These considerations motivate the following definition of state equivalence.

1 **Definition** Let $\boldsymbol{\alpha}_0{}^1$ and $\boldsymbol{\alpha}_0{}^2$ denote two states in Σ_{t_0}. Then $\boldsymbol{\alpha}_0{}^1$ and $\boldsymbol{\alpha}_0{}^2$ will be said to be *equivalent at* t_0, written as $\boldsymbol{\alpha}_0{}^1 \simeq \boldsymbol{\alpha}_0{}^2$, if and only if they are tags for the same aggregate, that is,

2 $$\mathcal{A}_{\boldsymbol{\alpha}_0{}^1}(t_0) = \mathcal{A}_{\boldsymbol{\alpha}_0{}^2}(t_0)$$

The conventional definition of equivalent states (see Ref. 1, p. 71) is somewhat less general than the above definition; it corresponds to making the following simplifying assumptions concerning $\mathcal{A}$.

Let $\mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$ denote the domain of the relation $\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)$, that is,

3 $$\mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)) \triangleq \{\mathbf{u}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$$

If $\mathcal{A}$ is such that $\mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$ is independent of both $\boldsymbol{\alpha}_0$ and t_0,† then $\mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$ will be denoted more simply by $\mathfrak{U}$ and will be called the *input-function space* of $\mathcal{A}$. Thus,

4 $$\mathfrak{U} \triangleq \{\mathbf{u}|(\mathbf{u},\mathbf{y}) \in \mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\} \qquad \text{independent of } \boldsymbol{\alpha}_0 \text{ and } t_0$$

If $\mathcal{A}$ has an input-function space, then the definition of equivalent states *2* can be replaced by the simpler definition

5 $$\{\boldsymbol{\alpha}^1 \simeq \boldsymbol{\alpha}^2\} \Leftrightarrow \{\forall \mathbf{u}[\bar{A}(\boldsymbol{\alpha}^1;\mathbf{u}) = \bar{A}(\boldsymbol{\alpha}^2;\mathbf{u})]\} \qquad \mathbf{u} \in \mathfrak{U}$$

where $\bar{A}(\boldsymbol{\alpha};\mathbf{u})$ and $\bar{A}(\boldsymbol{\alpha}^2;\mathbf{u})$ denote, respectively, the responses of $\mathcal{A}$ to $\mathbf{u}$ starting in states $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^2$. In words, *5* means that $\boldsymbol{\alpha}^1$ and $\boldsymbol{\alpha}^2$ are equivalent states of $\mathcal{A}$ if and only if for all inputs $\mathbf{u}$ in $\mathfrak{U}$ the response of $\mathcal{A}$ to $\mathbf{u}$ starting in state $\boldsymbol{\alpha}^2$ is identical with the response of $\mathcal{A}$ to $\mathbf{u}$ starting in state $\boldsymbol{\alpha}^1$.

The same concept of equivalence applies when we speak of a state $\boldsymbol{\alpha}$ of a system $\mathcal{A}$ as being equivalent to a state $\boldsymbol{\beta}$ of a system $\mathcal{B}$. In this case

6 $$\{\boldsymbol{\alpha} \simeq \boldsymbol{\beta}\} \Leftrightarrow \{\forall \mathbf{u}[\bar{\mathbf{A}}(\boldsymbol{\alpha};\mathbf{u}) = \bar{\mathbf{B}}(\boldsymbol{\beta};\mathbf{u})]\}$$

where $\bar{\mathbf{B}}[\boldsymbol{\beta};\mathbf{u}]$ denotes the response of $\mathcal{B}$ to $\mathbf{u}$ starting in state $\boldsymbol{\beta}$.

7 *Example* Consider the two systems of Example *1.3.3*, for which the respective input-output-state relations read

$$\mathcal{A}: \qquad y(t) = \alpha + \int_{t_0}^{t} u(\xi)\, d\xi$$

$$\mathcal{B}: \qquad y(t) = \beta_0 + \beta_1(t - t_0) + \int_{t_0}^{t} u(\xi)\, d\xi$$

† Note that $\mathfrak{D}(\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0))$ is independent of t_0 if $\mathcal{A}$ is time-invariant. However, this is not a necessary condition.

In this case, the state $\alpha = 0$ of $\mathcal{A}$ is equivalent to the state $\boldsymbol{\beta} = (0,0)$ $[\boldsymbol{\beta} \triangleq (\beta_0,\beta_1)]$ of $\mathcal{B}$.

8 *Example* Consider the finite-state systems $\mathcal{A}$ and $\mathcal{B}$ shown in Fig. *1.3.1*. Here it is obvious by inspection that $1 \simeq 1'$ and $2 \simeq 2'$. On the other hand, the state $3'$ of $\mathcal{B}$ has no equivalent in $\mathcal{A}$.

9 **Strong equivalence** In Sec. *3* two systems $\mathcal{A}$ and $\mathcal{B}$ were defined to be equivalent, written as $\mathcal{A} = \mathcal{B}$, if $\mathcal{A} = \mathcal{B}$ in the set-theoretic sense, that is, if every input-output pair belonging to $\mathcal{A}$ also belongs to $\mathcal{B}$ and vice versa. As was pointed out, if $\mathcal{A} = \mathcal{B}$, then $\mathcal{A}$ and $\mathcal{B}$ are indistinguishable by a simple experiment.

There are many cases in which a stronger concept of equivalence, namely one in which $\mathcal{A}$ and $\mathcal{B}$ are indistinguishable by any multiple experiment, is more appropriate. A convenient way of expressing strong equivalence of $\mathcal{A}$ and $\mathcal{B}$ is contained in the following definition.

10 **Definition** $\mathcal{A}$ and $\mathcal{B}$ are *strongly equivalent,* written as $\mathcal{A} \equiv \mathcal{B}$, if and only if, for each t_0, to every state $\boldsymbol{\alpha}$ in the state space of $\mathcal{A}$ there is an equivalent state $\boldsymbol{\beta}$ in the state space of $\mathcal{B}$ and vice versa. In symbols, if the responses of $\mathcal{A}$ and $\mathcal{B}$ to an input $\mathbf{u}$ starting in states $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ at time t_0 are denoted by $\bar{A}(\boldsymbol{\alpha};\mathbf{u})$ and $\bar{B}(\boldsymbol{\beta};\mathbf{u})$, respectively, then we can write compactly

11 $$\mathcal{A} \equiv \mathcal{B} \Leftrightarrow \forall\boldsymbol{\alpha}\exists\boldsymbol{\beta}\forall\mathbf{u}\forall t_0[\bar{A}(\boldsymbol{\alpha};\mathbf{u}) = \bar{B}(\boldsymbol{\beta};\mathbf{u})]$$

and

$$\forall\boldsymbol{\beta}\exists\boldsymbol{\alpha}\forall\mathbf{u}\forall t_0[\bar{A}(\boldsymbol{\alpha};\mathbf{u}) = \bar{B}(\boldsymbol{\beta};\mathbf{u})]$$

In similar symbols, the expression for the definition of equivalence becomes

12 $$\mathcal{A} = \mathcal{B} \Leftrightarrow \forall\boldsymbol{\alpha}\forall\mathbf{u}\forall t_0\exists\boldsymbol{\beta}[\bar{A}(\boldsymbol{\alpha};\mathbf{u}) = \bar{B}(\boldsymbol{\beta};\mathbf{u})]$$

and

$$\forall\boldsymbol{\beta}\forall\mathbf{u}\forall t_0\exists\boldsymbol{\alpha}[\bar{A}(\boldsymbol{\alpha};\mathbf{u}) = \bar{B}(\boldsymbol{\beta};\mathbf{u})]$$

Note that the only difference between *11* and *12* stems from the difference in the orders of quantifiers. Thus, in the case of equivalence, the state $\boldsymbol{\beta}$ in *12* depends on $\boldsymbol{\alpha}$, $\mathbf{u}$, and t_0, whereas in the case of strong equivalence, $\boldsymbol{\beta}$ in *11* depends only on $\boldsymbol{\alpha}$.

The notions of equivalence and strong equivalence are of central importance in system theory. A detailed discussion of the properties and consequences of these notions is presented in [1]. Here we shall restrict ourselves to a few comments centering on the role of the notion of state in system equivalence.

First, we observe that the notion of equivalence is independent of the notion of state whereas that of strong equivalence is not. This implies that the equivalence of $\mathcal{A}$ and $\mathcal{B}$ does not imply that $\mathcal{A}$ and $\mathcal{B}$ are strongly equivalent, although the reverse is clearly true, that is, $\mathcal{A} \equiv \mathcal{B} \Rightarrow \mathcal{A} = \mathcal{B}$. Essentially, this means that the indistinguishability of $\mathcal{A}$ and $\mathcal{B}$ under a single experiment does not imply their indistinguishability under a multiple experiment.

To illustrate, consider the finite-states system shown in Fig. *1.3.2*. It is easy to verify by inspection that every input-output pair of $\mathcal{A}$ is also an input-output pair for $\mathcal{B}$, and vice versa. On the other hand, the state $3'$ of $\mathcal{B}$ is clearly not equivalent to any state of $\mathcal{A}$. Hence, $\mathcal{A}$ and $\mathcal{B}$ are not strongly equivalent.

While it is not true in general that $\mathcal{A} = \mathcal{B} \Rightarrow \mathcal{A} \equiv \mathcal{B}$, there are types of systems for which equivalence implies strong equivalence. Among such systems [1] are strongly connected finite-state systems and differential systems of finite order. Note that the system $\mathcal{B}$ in Fig. *1.3.2* is not strongly connected, since state $3'$ in $\mathcal{B}$ is not reachable from other states.

7 *Association of states with a system*

In this section, we shall focus our attention on the following basic problem: Suppose that a system $\mathcal{A}$ is defined as a set of input-output pairs,

1 $$\mathcal{A} = \{(\mathbf{u},\mathbf{y})\}$$

possibly, but not necessarily, through an input-output relation. If we can find a family of bundles $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}_{\boldsymbol{\alpha}_0,t_0}$ satisfying the conditions of covering, closure under left truncation, uniqueness, and continuation, then, as shown in Sec. *5*, the tags $\{\boldsymbol{\alpha}_0\}$ of aggregates $\{\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0)\}$ are the states of $\mathcal{A}$ and the relation

2 $$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\alpha}_0;\mathbf{u}_{[t_0,t]})$$

is an input-output-state relation for $\mathcal{A}$. The problem is: How can one find a family of bundles satisfying these conditions, i.e., a family of aggregates of $\mathcal{A}$? Since the states of $\mathcal{A}$ are the tags of its aggregates, we shall refer to this problem as the problem of *associating states with a system.*

The association of states with a system $\mathcal{A}$ which is defined as a collection of input-output pairs results in a *system with a state structure*, or SSS, for short, which is characterized by the input-output-state relation *2* and which is equivalent to $\mathcal{A}$ in the sense of *1.3.8*. It is important to note, however, that if $\mathcal{A}_\sigma$ is an SSS which corresponds to a particular aggregation of input-output pairs in $\mathcal{A}$ and if $\mathcal{A}_\lambda$ is another such SSS which corresponds to a different mode of aggregation of input-output pairs in $\mathcal{A}$, then, in general, $\mathcal{A}_\sigma$ and $\mathcal{A}_\lambda$ will not be strongly equivalent in the sense of *1.6.10*. Clearly, if all the SSS which are equivalent to $\mathcal{A}$ are strongly equivalent to one another, then there is one and only one family of bundles which qualify as aggregates of $\mathcal{A}$.

As an aid in finding a family of aggregates of $\mathcal{A}$, it is very helpful to establish a basic property of input-output-state relations which, as will be seen below, is equivalent to the continuation condition. Specifically, from inspection of *1.2.13*, it is obvious that, since the input-output pair $(\mathbf{u}^1,\mathbf{y}^1)$ belongs to the bundle $\mathcal{A}_{\boldsymbol{\alpha}_1}(t_1)$, we can write

3 $$\mathbf{y}^1(t) = \mathbf{A}(\boldsymbol{\alpha}_1;\mathbf{u}^1_{[t_1,t]})$$

where $\boldsymbol{\alpha}_1$ is independent of $\mathbf{u}^1$ and is dependent solely on $\boldsymbol{\alpha}_0$ and $\mathbf{u}^0$, that is,

4 $$\boldsymbol{\alpha}_1 = \mathbf{F}(\boldsymbol{\alpha}_0;\mathbf{u}^0_{[t_0,t_1]})$$

Furthermore, it is also clear that the continuation condition is satisfied if *3* and *4* hold true.

This observation makes it possible to make the following assertion: Suppose that the bundling of input-output pairs of $\mathcal{A}$ is defined by the input-output-state relation

5 $$\mathbf{y}(t) = \mathbf{A}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]})$$

through

6 $$\mathcal{A}_{\boldsymbol{\alpha}_0}(t_0) \triangleq \{(\mathbf{u}_{[t_0,t]},\bar{\mathbf{A}}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}))\}_{\mathbf{u}}$$

where $\boldsymbol{\alpha}_0 = \mathbf{x}(t_0) =$ initial state, $\mathbf{u} = \mathbf{u}_{[t_0,t]}$, and

7 $$\mathbf{y}_{[t_0,t]} = \bar{\mathbf{A}}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]})$$

is an input-output-state relation (equivalent to *5*) which expresses the output segment [rather than $\mathbf{y}(t)$] as a function of the initial state $\mathbf{x}(t_0)$ and the input segment $\mathbf{u}_{[t_0,t]}$. Then the bundling in question satisfies the continuation condition if and only if the input-output-state relation *5* satisfies the identity

8 $$\mathbf{y}(t) = \mathbf{A}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}) = \mathbf{A}(\mathbf{x}(t_1);\mathbf{u}_{[t_1,t]})$$

where $\mathbf{x}(t_1)$ depends only on $\mathbf{x}(t_0)$ and $\mathbf{u}_{[t_0,t_1]}$, that is,

9 $$\mathbf{x}(t_1) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t_1]})$$

This identity expresses the so-called *response separation property* of *5*. Thus, the continuation condition implies and is implied by the response separation property.

In effect, the response separation property of the input-output-state relation affords a convenient way of checking on whether or not the continuation condition is satisfied by the family of bundles which is generated by the input-output-state relation in question. Using this property, the association of states with a given system involves only two steps: (*a*) Finding a relation of the form *2*, with $\boldsymbol{\alpha}_0$ ranging over a space Σ_{t_0}, such that every input-output pair in $\mathcal{A}$ can be represented in this form and, conversely, for every t_0 and every $\boldsymbol{\alpha}_0$ in Σ_{t_0}, $(\mathbf{u},\bar{\mathbf{A}}(\mathbf{x}(t_0);\mathbf{u}))$ is an input-output pair in $\mathcal{A}$; and (*b*) verifying that *2* has the response separation property. If (*a*) and (*b*) are satisfied, then the $\boldsymbol{\alpha}_0$ qualify as the states of $\mathcal{A}$ at time t_0 and the relation

10 $$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\alpha}_0;\mathbf{u})$$

may be regarded as an input-output-state relation for $\mathcal{A}$.

As was pointed out earlier (Sec. *5*), when $\mathcal{A}$ is defined by an input-output relation which has the form of a differential or difference equation, a relation of the form *2* constitutes an expression for the general solution of the equation. In such cases, then, the association of states with a system involves merely a check on whether the general solution has the response separation property.

11 *Example* As a simple illustration of the above procedure, suppose that $\mathcal{A}$ is defined by the input-output relation

12 $$(p + 1)y = u$$

In this case, the general solution may be written as

13 $$y(t) = \alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)}u(\xi)\, d\xi \qquad t \geq t_0$$

which is of the form $y(t) = A(\alpha_0;u_{[t_0,t]})$, with α_0 ranging over the real line $(-\infty,\infty)$.

To verify that *13* has the response separation property, we have to verify that there exists a real number α_1 independent of $u_{[t_1,t]}$ such that, for all α_0, t_0, t_1, t, and u, we have the identity

14 $$\alpha_0 e^{-(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)}u(\xi)\, d\xi = \alpha_1 e^{-(t-t_1)} + \int_{t_1}^{t} e^{-(t-\xi)}u(\xi)\, d\xi$$

Clearly, *14* can be satisfied by relating α_1 to α_0 and $u_{[t_0,t_1]}$ by the equation

15 $$\alpha_1 = \alpha_0 e^{-(t_1-t_0)} + \int_{t_0}^{t_1} e^{-(t_1-\xi)}u(\xi)\, d\xi$$

This equation is a concrete form of *4* for the case under consideration. In effect, it is a state equation for $\mathcal{A}$ induced by the input-output-state relation *13*.

It is of interest to observe that *13* would not have the response separation property if the exponents of the two terms in the right-hand member of *13* were not identical. For example, if $y(t)$ were given, say, as

16 $$y(t) = \alpha_0 e^{-2(t-t_0)} + \int_{t_0}^{t} e^{-(t-\xi)} u(\xi)\, d\xi$$

then *16* would not have the response separation property. More generally, it is easy to verify that if the input-output-state relation is of the form

17 $$y(t) = \sum_{i=1}^{n} \alpha_i e^{-a_i(t-t_0)} + \int_{t_0}^{t} h(t-\xi) u(\xi)\, d\xi$$

where the α_i range over the real line, then the response separation property requires that the function $h(t)$ be a linear combination of the exponentials $e^{-a_i t}$, $i = 1, \ldots, n$.

State separation property

In the foregoing discussion, we assumed that the point of departure for associating states with a given system $\mathfrak{A}$ is an input-output-state relation for $\mathfrak{A}$ of the form

18 $$\mathbf{y}(t) = \mathbf{A}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}) \qquad t \geq t$$

where $\mathbf{x}(t_0)$, the state at time t_0, is a variable ranging over a state space Σ_{t_0}.

If *18* has the response separation property, then we can write

19 $$\mathbf{A}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}) = \mathbf{A}(\mathbf{x}(t_1);\mathbf{u}_{[t_1,t]})$$

where $\mathbf{x}(t_1)$ is related to $\mathbf{x}(t_0)$ and $\mathbf{u}_{[t_0,t]}$ by the state equation

20 $$\mathbf{x}(t_1) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t_1]})$$

As shown in [1], an important consequence of the response separation property is the so-called *state separation property* (or the *semigroup property*) of the state equation *20*. When written in a form analogous to *19*, this property may be expressed as the identity

21 $$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]}) = \mathbf{F}(\mathbf{x}(t_1);\mathbf{u}_{[t_1,t]})$$

where

22 $$\mathbf{x}(t_1) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t_1]})$$

with the understanding that *21* holds for all t_0, t_1, t $(t_0 \leq t_1 \leq t)$, $\mathbf{x}(t_0)$, and $\mathbf{u}_{[t_0,t]}$. In words, the state separation property means: If an input segment $\mathbf{u}_{[t_0,t]}$ is divided arbitrarily into two segments $\mathbf{u}_{[t_0,t_1]}$ and $\mathbf{u}_{[t_1,t]}$, and if an initial state $\mathbf{x}(t_0)$ is carried by $\mathbf{u}_{[t_0,t]}$ into $\mathbf{x}(t)$ and by $\mathbf{u}_{[t_0,t_1]}$ into $\mathbf{x}(t_1)$, then $\mathbf{x}(t_1)$ is carried by $\mathbf{u}_{[t_1,t]}$ into $\mathbf{x}(t)$.

In most cases of practical interest, the dependence of $\mathbf{y}(t)$ on $\mathbf{u}_{[t_0,t]}$

in the input-output-state relation *18* is such that, as $t_0 \rightarrow t$, *18* tends to a relation of the form

23 $$\mathbf{y}(t) = g(\mathbf{x}(t),\mathbf{u}(t),t)$$

which implies that the output at time t depends only on the state at time t, the input at time t, and t. Systems of this type will be said to be *proper*. An example of a proper system (see Ref. 1, chap. 4) is a system defined by an input-output relation of the form

24 $$(a_n p^n + \cdots + a_0)y = (b_m p^m + \cdots + b_0)u$$

in which $n \geq m$.

A proper system may be characterized by the equation

25 $$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t),\mathbf{u}(t),t)$$

and the state equation

26 $$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]})$$

In effect, these two equations define the system in question through the input-output-state relation

27 $$\mathbf{y}(t) = \mathbf{g}(\mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]},\mathbf{u}(t),t))$$

which results from substituting *26* into *25*.

An important property of proper systems may be expressed in the following form.

28 **Assertion** If the state equation *26* has the state separation property, then the input-output-state relation *27* will have the response separation property and, consequently, the $\mathbf{x}(t)$ in *26* will qualify as the states of the system defined by *25* and *26*.

The truth of this assertion follows at once from making use of *21* and *22* in *27*.

The above assertion provides a very convenient way of associating states with a *differential system*, that is, a system whose input-output relation has the form of a differential equation. Thus, suppose that the given differential equation (or equations) defining $\mathcal{A}$ can be shown to be equivalent to the equations

29 $$\mathbf{y}(t) = \mathbf{g}(\mathbf{x}(t),\mathbf{u}(t),t)$$

30 $$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t),t)$$

in which $\mathbf{x}(t)$ ranges over a space Σ_t. Then, the following can be asserted (corollary 2.3.36 in Ref. 1):

31 **Assertion** If equation *30* has a unique solution for $\mathbf{x}(t)$ for every initial state $\mathbf{x}(t_0)$, then the solution, expressed as

32 $$\mathbf{x}(t) = \mathbf{F}(\mathbf{x}(t_0);\mathbf{u}_{[t_0,t]})$$

will have the state separation property. Coupled with Assertion *28*, this implies that if we can demonstrate that *29* and *30* are equivalent to the differential equation (or equations) defining $\mathcal{A}$ (in the sense of having the same set of solutions), then we can conclude that the $\mathbf{x}(t)$ in *29* and *30* qualify as the states of $\mathcal{A}$.

As a simple illustration, consider a system $\mathcal{A}$ defined by the input-output relation

33 $$(a_n p^n + \cdots + a_0)y = u$$

in which the a_i are real-valued coefficients and the input and output at time t are real-valued variables. On defining

34 $$\begin{aligned} x_1(t) &= y(t) \\ x_2(t) &= py(t) \\ &\cdots\cdots\cdots \\ x_n(t) &= p^{n-1}y(t) \end{aligned}$$

the differential equation *31* can be replaced by the single vector differential equation

35 $$\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}u(t)$$

together with the equation

36 $$y(t) = \mathbf{C}\mathbf{x}(t)$$

in which $\mathbf{x} = (x_1, \ldots, x_n)$ and the matrices $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{C}$ are given by

37 $$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ -\dfrac{a_0}{a_n} & \cdot & \cdot & \cdots & -\dfrac{a_{n-1}}{a_n} \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \dfrac{1}{a_n} \end{bmatrix}$$

$$\mathbf{C} = [1 \quad 0 \quad \cdots \quad 0]$$

Equation *35* has a unique solution which can be written as

38 $$\mathbf{x}(t) = \mathbf{\Phi}(t - t_0)\mathbf{x}(t_0) + \int_{t_0}^{t} \mathbf{\Phi}(t - \xi)\mathbf{B}u(\xi)\, d\xi$$

where $\mathbf{\Phi}(t)$ is the solution of the equation

39 $$\dot{\mathbf{\Phi}} = \mathbf{A\Phi} \qquad \mathbf{\Phi}(0) = \mathbf{I} \qquad \text{identity matrix}$$

It follows from Assertion *30*—and can also be readily verified directly—that *38* has the state separation property. Consequently, from *28* we can conclude that $\mathbf{x}(t)$ as defined by *34* qualifies as a state vector for $\mathcal{A}$.

In more general cases, the association of states with a system can be carried out in a similar manner, with the response and state separation properties of the input-output-state relations and state equations serving as the bridge between the definition of a state as a tag for an aggregate and its expression in terms of the input and output of $\mathcal{A}$. Detailed expositions of the techniques for associating states with a system may be found in some of the references listed below, especially [1] and [25]. In addition, a number of methods relevant to this problem in the context of linear differential systems will be discussed in Chap. *5*.

REFERENCES

1 Zadeh, L. A., and C. A. Desoer: "Linear System Theory," McGraw-Hill, New York, 1963.
2 Zadeh, L. A.: The Concept of State in System Theory, pp. 39–50, in "Views on General Systems Theory," M. Mesarovic (ed.), Wiley, New York, 1964.
3 Mesarovic, M.: Foundations for a General Systems Theory, pp. 1–24, in "Views on General Systems Theory," M. Mesarovic (ed.), Wiley, New York, 1964.
4 Arbib, M.: Automata Theory and Control Theory—A Rapprochement, *Automatika*, vol. 3, pp. 161–189, 1966.
5 Kalman, R.: Algebraic Aspects of the Theory of Dynamical Systems, pp. 133–146, in *Proc. Intern. Symp. Differential Equations and Dynamical Systems,* J. Hale and J. LaSalle (eds.), Academic, New York, 1967.
6 Wymore, W.: "A Mathematical Theory of Systems Engineering," Wiley, New York, 1967.
7 Salovaara, S.: On Set Theoretical Foundations of System Theory—A Study of the State Concept, *Acta Polytech. Scand.*, no. 15, pp. 1–74, 1967.
8 Nerode, A.: Linear Automaton Transformations, *Proc. Am. Math. Soc.*, vol. 9, pp. 541–544, 1958.
9 Windeknecht, T.: Mathematical Systems Theory: Causality, *Math. Systems Theory,* vol. 1, no. 4, pp. 279–288, 1967.
10 Krohn, K., and J. Rhodes: Algebraic Theory of Machines, pp. 341–384, in *Proc. Symp. Math. Theory of Automata,* J. Fox (ed.), *Polytech. Inst. Brooklyn,* New York, 1962.
11 Krohn, K., and J. Rhodes: Algebraic Theory of Machines, *Trans. Am. Math. Soc.*, vol. 115, pp. 450–464, 1965.

12 Zeiger, P.: Cascade Decomposition of Automata Using Covers, in M. Arbib (ed.), "Algebraic Theory of Machines, Languages and Semigroups," pp. 55–80, Academic, New York, 1968.
13 Kalman, R. E.: On the General Theory of Control Systems, *Proc. First Intern. Congr. IFAC, Moscow*, 1960; also "Automatic and Remote Control," pp. 481–492, Butterworth, London, 1961.
14 Ogata, K.: "State Space Analysis of Control Systems," Prentice-Hall, Englewood Cliffs, N.J., 1966.
15 Newcomb, R. W.: "Linear Multiport Synthesis," McGraw-Hill, New York, 1966.
16 Kuh, E. S., and R. A. Rohrer: "Theory of Linear Active Networks," Holden-Day, San Francisco, 1967.
17 Fel'dbaum, A. A.: "Optimal Control Systems," Academic, New York, 1965.
18 Athans, M., and P. L. Falb: "Optimal Control," McGraw-Hill, New York, 1966.
19 Schwarz, R. J., and B. Friedland: "Linear Systems," McGraw-Hill, New York, 1965.
20 Tou, J. T.: "Modern Control Theory," McGraw-Hill, New York, 1964.
21 Timothy, L. K., and B. E. Bona: "State Space Analysis," McGraw-Hill, New York, 1968.
22 Peschon, J.: "Discipline and Techniques of System Synthesis," Blaisdell, Boston, 1965.
23 Leondes, C.: "Modern Control Systems Theory," McGraw-Hill, New York, 1965.
24 Zadeh, L. A.: "Fuzzy Sets and Systems," *Proc. Symp. System Theory, Polytech. Inst. Brooklyn*, April, 1965, pp. 29–37.
25 Ho, B. L., and R. Kalman: Effective Construction of State Variable Models from Input-output Relations, pp. 449–459, in *Proc. 3d Allerton Conf.*, 1965.
26 Arbib, M.: A Common Framework for Automata Theory and Control Theory, *SIAM J. Control*, vol. 3, no. 2, pp. 206–222, 1965.